Cue validity learning in threat classification judgments

David J. Bryant

Defence R&D Canada – Toronto

Technical Report

DRDC Toronto TR 2003-041

March 2003

Author

David J. Bryant, Ph.D.

Approved by

Ross Pigeau

Head, Command Effectiveness and Behaviour Section

Approved for release by

K.M. Sutton

Chair, Document Review and Library Committee

© Her Majesty the Queen as represented by the Minister of National Defence, 2003

© Sa majesté la reine, représentée par le ministre de la Défense nationale, 2003

104-05561/

Abstract

This report describes an experiment that investigated probabilistic cue learning in a simulated naval warfare threat classification task. The *Fast and Frugal Heuristic* approach was employed to develop an heuristic, Called the "Take-the-Best-for-Classification" (TTB-C) heuristic, that performs the threat classification task with minimal information and computation. Two variables were manipulated in this experiment. The first, varied between subjects, was the Instruction Set given to participants (Describe vs. Discriminate), which emphasized either the patterns of cue values associated with friend and foe contacts or the differences in typical cue patterns between the two types of contact. The second variable, varied within subjects, was the size of the differences among cue validities (Cue Validity Differences) of the four cues. Four hypotheses were derived from the TTB-C heuristic and tested. Although the results provided support for only one hypothesis, further studies are warranted to explore the potential use of fast and frugal heuristics under conditions of uncertainty, time pressure, and resource costs imposed on data gathering.

Résumé

Ce rapport décrit une expérience visant à étudier l'apprentissage de repères probabilistes dans une fonction de classification des dangers d'une guerre navale simulée. L'approche *heuristique simple et rapide* a été utilisée pour élaborer une heuristique, appelée « ne garder que le meilleur en vue de la classification » (TTB-C) qui remplit la fonction de classification des dangers avec un minimum d'information et de calculs. On a manipulé deux variables au cours de cette expérience. La première, qui variait d'un sujet à l'autre, était le jeu d'instructions remis aux participants (Décrire par opposition à Distinguer), qui mettait l'accent soit sur les modèles de valeurs des repères associées aux contacts amis ou ennemis, soit sur les différences entre deux sortes de contact dans les modèles de repères types. La seconde variable, qui variait à l'intérieur des sujets, était l'importance des différences entre les validités des repères (différences de validité des repères) des quatre repères. On a tiré quatre hypothèses de l'heuristique TTB-C et on les a testées. Bien que les résultats n'appuient qu'une hypothèse, il faut faire d'avantage d'études pour explorer l'usage qu'on pourrait faire des heuristiques simples et rapides quand l'incertitude règne, le temps presse et le coût des ressources influe sur la collecte de données.

This page intentionally left blank.

Executive summary

This report describes an experiment that investigates how people use different pieces of information (cues) to classify contacts (threat classification task) in a simulated naval warfare environment. This topic is of importance to situation and threat assessment because characteristics associated with sensor contacts are variable and friendly, hostile, and neutral platforms can possess similar characteristics under a range of conditions. Such characteristics can be termed "probabilistic cues" because they are associated with a kind of contact at some level of probability or chance.

An important factor explored in past research on probabilistic cue learning has been *cue validity*, which is the extent to which a cue correctly indicates the state of a criterion variable (validity is typically defined in terms of Bayesian probabilities or the regression of one variable on another). Numerous studies have demonstrated that people are able to learn the validities of cues and use them in predicting outcomes but that this is often difficult and a number of factors can negatively affect cue learning.

The *Fast and Frugal Heuristic* approach offers a theoretical perspective on decision making based on a conceptualization of rationality in which behaviour is evaluated in terms of its adaptiveness within the limits of time and knowledge imposed by the situation and the computational power and the decision maker [1]. Designed to be effective and psychologically plausible, fast and frugal heuristics offer a way to develop detailed computational models of the threat classification task. This report proposes a fast and frugal heuristic that performs the threat classification task with minimal information and computation. Called the "Take-the-Best-for-Classification" (TTB-C) heuristic, it is based on an established heuristic developed to perform a two-alternative choice task [2]. TTB-C seeks the most valid cue to threat class available and assigns a threat classification based on the value of that cue alone. This contrasts with more complex methods that weigh all available information.

To assess the extent to which participants' cue selection and classification behaviour would be consistent with the TTB-C heuristic, an experiment was devised to test the following hypotheses concerning peoples' threat classification performance:

1. When provided with trial-and-error learning experience, participants will learn the relative ranking of cues according to their validity in predicting threat class;

2. During a test session in which participants must select a cue to inspect its value, participants will select the most valid cue first for virtually all items;

3. Participants will rarely select any additional cues because the most valid cue will always be available and TTB-C can make a decision without additional information (additional cue selections would likely reflect attention lapses and response errors); and

4. Participants' accuracy rate in judgments of threat class will be roughly 90%, the proportion expected if they exclusively use only the most predictive cue in this experiment.

Twenty-four men and women with no naval command experience performed a simulated threat classification task on the Team and Individual Threat Assessment Network (TITAN) platform. The TITAN interface presented participants with a radar screen on which "contacts," each corresponding to a single entity around the participant's "own ship," were presented by asterisk symbols. Using the computer mouse, the participants clicked on ("hooked") a contact, which activated a set of buttons that allowed access to cues for that contact. By clicking another button, participants called up a box in which they selected one of two possible classifications, "friend" or "foe." An additional windows then opened and participants indicated a confidence judgment for their classification and received feedback concerning their classification accuracy. Participants classified 200 contacts (100 friend and 100 foe) in a training session, during which they received accuracy feedback, and 100 contacts (50 friend and 50 foe) in a test session, during which they received no feedback. Following the training and test phases, subjects were asked to indicate the validities of each cue as a predictor of contact classification.

Two variables were manipulated in this experiment. The first, varied between subjects, was the Instruction Set given to participants (Describe vs. Discriminate), which emphasized either the patterns of cue values associated with friend and foe contacts or the differences in typical cue patterns between the two types of contact, respectively. The second variable, varied within subjects, was the size of the differences among cue validities (Cue Validity Differences) of the four cues.

Overall, the results of the experiment provided no evidence consistent with any hypothesis other than the first. Participants did seem to learn the relative validities of the four cues, as evidenced by their estimates of the probabilistic relations of each cue to threat class, although they did not necessarily learn a completely accurate ranking of cue validities (Hypothesis 1). The remaining hypotheses were clearly falsified. Participants' average accuracy rates were well below the 90% level they could have achieved by following TTB-C (Hypothesis 4). Examination of participants' cue selection patterns revealed that, first, participants generally did not select the most valid cue first more often than would be expected by chance (Hypothesis 2), and second, participants selected more than one cue (and often all four cues) for inspection more often than would be expected due to error or attention lapses (Hypothesis 3). Thus, participants were neither frugal nor systematic in their cue selection, aside from one participant who's behaviour did conform Hypotheses 2 and 3 in one condition.

Although participants did not seem to employ the TTB-C heuristic, it remains a plausible strategy for threat assessment because it offers the decision maker great cognitive economy and allows a fairly high level of performance in a task environment in which a few cues are highly predictive. Future studies will examine the potential use of fast and frugal heuristics in threat classification tasks under conditions of limited information, time pressure, and resource costs imposed on data gathering.

Bryant, D.J. 2003. Cue validity Learning in threat classification judgments. DRDC Toronto TR 2003-041. Defence R&D Canada – Toronto.

Sommaire

Ce rapport décrit une expérience visant à voir comment les gens utilisent différents éléments d'information (repères) pour classifier des contacts (fonction de classification des dangers) dans un milieu de combat naval simulé. C'est un sujet important pour l'évaluation de la situation et des dangers parce que les caractéristiques reliées aux contacts de senseurs sont variables, et que des plates-formes amies, hostiles et neutres peuvent posséder des caractéristiques semblables dans diverses conditions. On peut dire que de telles caractéristiques sont des « repères probabilistes » parce qu'on les associe avec une sorte de contact à un niveau quelconque de probabilité ou de hasard.

Au cours de recherches antérieures sur l'apprentissage de repères probabilistes, on a étudié un facteur important, à savoir, la *validité du repère*, qui est la mesure dans laquelle un repère indique correctement l'état d'un critère variable (la validité est généralement définie par rapport aux probabilités bayésiennes ou à la régression d'une variable sur une autre). De nombreuses études ont montré que les gens peuvent apprendre les validités de repères et les utiliser pour prédire des résultats, mais c'est souvent difficile, et un nombre de facteurs peuvent influer négativement sur l'apprentissage de repères.

L'approche *heuristique simple et rapide* offre une perspective théorique de la prise de décision basée sur une conceptualisation de la rationalité dans laquelle le comportement est évalué par rapport à son adaptivité dans les limites de temps et de connaissance qu'imposent la situation, la capacité de traitement et le décideur [1]. Conçue pour être efficace et plausible sur le plan psychologique, l'heuristique simple et rapide offre un moyen d'élaborer des modèles informatiques détaillés de la fonction de classification des dangers. Le présent rapport propose une heuristique simple et rapide qui remplit la fonction de classification des dangers avec un minimum d'information et de calcul. L'heuristique, qui consiste à « ne garder que le meilleur en vue de la classification » (TTB-C), repose sur une heuristique établie élaborée pour remplir une fonction de choix à partir d'une alternative [2]. TTB-C cherche le repère le plus valide de la classe de danger disponible et attribue une classification des dangers en fonction de la valeur du repère uniquement. Ceci diffère de méthodes plus complexes qui tiennent compte de toute l'information disponible.

Pour évaluer la mesure dans laquelle les repères et la classification que choisissent les participants correspondraient à l'heuristique TTB-C, on a mis au point une expérience visant à vérifier les hypothèses suivantes relatives à la classification des dangers que les gens choisissent :

1. Quand les participants apprennent par essais et erreurs, ils apprennent le classement relatif des repères en fonction de leur validité servant à prédire la classe de danger;

2. Au cours d'une séance d'essai durant laquelle ils doivent sélectionner un repère pour en étudier la valeur, les participants choissent d'abord le repère le plus valide pour pratiquement tous les points;

3. Les participants choisissent rarement des repères supplémentaires parce que le repère le plus valide est toujours disponible et que TTB-C peut prendre une décision sans information supplémentaire (le fait de choisir des repères supplémentaires proviendrait probablement d'un relâchement de l'attention et d'erreurs dans les réponses); et

4. Le taux de précision des jugements pour la classe de danger est d'environ 90 %, la proportion à laquelle on s'attend si les participants utilisent exclusivement le repère le plus prédictif pendant cette expérience.

Vingt-quatre hommes et femmes sans aucune expérience du commandement naval ont rempli une fonction de classification des dangers selon la plate-forme du Réseau d'évaluation des menaces pour l'individu ou le groupe (plate-forme du TITAN). L'interface TITAN montrait aux participants un écran radar sur lequel des « contacts », chacun correspondant à une seule entité dans les parages du « vaisseau » du participant, étaient représentés par des astérisques. Avec la souris de l'ordinateur, les participants ont cliqué sur un contact, ce qui activait un jeu de boutons donnant accès à des repères pour ce contact. En cliquant sur un autre bouton, les participants amenaient à l'écran une boîte dans laquelle ils choisissaient une des deux classifications possibles, « ami » ou « ennemi ». Une autre fenêtre s'ouvrait alors, et les participants indiquaient un jugement de confiance envers la classification qu'ils avaient choisie et recevaient une réaction sur la précision de cette dernière. Les participants ont classifié 200 contacts (100 amis et 100 ennemis) pendant une séance d'entraînement, durant laquelle ils ont reçu de la réaction, et 100 contacts (50 amis et 50 ennemis) pendant une séance d'essai durant laquelle ils n'ont pas reçu de réaction. Après les séances d'entraînement et d'essai, on a demandé aux participants d'indiquer les validités de chaque repère en tant que variable prédictive de la classification du contact.

Deux variables ont été manipulées pendant cette expérience. La première, qui variait d'un sujet à l'autre, était le jeu d'instructions remis aux participants (Décrire par opposition à Distinguer), qui mettait l'accent sur les modèles de valeurs des repères associées aux contacts amis ou ennemis, ou sur les différences, dans des modèles de repères types, entre les deux sortes de contact, respectivement. La seconde variable, qui variait à l'intérieur des sujets, était l'importance des différences entre les validités des repères (Différences de validité des repères) des quatre repères.

Dans l'ensemble, les résultats de l'expérience n'ont pas fourni d'évidence confirmant les hypothèses, sauf la première. Les participants ne semblaient pas apprendre les validités relatives des quatre repères, comme le montrent leurs estimations des relations de probabilité entre chaque repère et la classe de danger, bien qu'ils n'aient pas appris nécessairement un classement complet et précis des validités des repères (Hypothèse 1). Les autres hypothèses étaient clairement falsifiées. Les taux moyens de précision des participants étaient nettement inférieurs aux 90 % qu'ils auraient pu obtenir en suivant TTB-C (Hypothèse 4). L'examen des modèles de sélection des repères a révélé que, premièrement, les participants ne choisissaient généralement pas en premier le repère le plus valide, plus souvent qu'on ne s'y attendrait en fonction du hasard (Hypothèse 2) et, deuxièmement, les participants choisissaient plus d'un repère (et souvent les quatre repères) pour l'inspection, plus souvent qu'on ne s'y attendrait en fonction des erreurs et du manque d'attention (Hypothèse 3). Par conséquent, les participants n'étaient ni simples, ni systématiques quand ils choisissaient leurs repères, à l'exception d'un d'entre eux dont le comportement correspondait aux Hypothèses 2 et 3 pour une condition.

Les participants ne semblaient pas utiliser l'heuristique TTB-C, mais elle demeure néanmoins une stratégie plausible d'évaluation des dangers parce qu'elle donne au décideur une plus grande économie cognitive et permet un niveau de rendement assez élevé dans un champ d'intervention offrant peu de repères très prédictifs. D'autres études serviront à examiner l'utilisation éventuelle de l'heuristique rapide et simple dans les fonctions de classification des dangers quand l'information est limitée, le temps presse et le coût des ressources influe sur la collecte de données.

Bryant, D.J. 2003. Cue validity Learning in threat classification judgments. DRDC Toronto TR 2003-041. Defence R&D Canada – Toronto.

This page intentionally left blank.

Table of contents

Abstract i

Ré.umé i

Executive summary iii

Sommaire v

Table of contents ix

List of figures xi

List of tables xi

Acknowledgements xii

Introduction 1

- Background 1
- Probabilistic cue learning 2
- Fast and frugal heuristics 3
- The Take-the-Best heuristic 5
- The Take-the-Best-for-Classification heuristic 6
- Purpose of study 8

Method 10

- Participants 10
- Materials 10
- Design 12
- Procedure 13

Results 15

- Training session 15
- Test session 18
- Cue selections 24

Discussion 28
Evaluation of TTB-C 28
Other decision rules 30
The relevance of fast and frugal heuristics to issues of Command and Control 33

References 34

Annex A 38
Learning Probabilistic Cues Instructions (Discriminate Condition) 38
Learning Probabilistic Cues Instructions (Describe Condition) 43

List of symbols/abbreviations/acronyms/initialisms 48

List of figures

Figure 1. The Take-the-Best-for-Classification (TTB-C) Heuristic 7

Figure 2. TITAN interface (shows the cue selection screen; separate windows appear for threat classification judgments and confidence rating) 11

Figure 3. Classification Accuracy by Block in the Training Session 16

Figure 4. Average Number of Cues Selected by Sequence 25

Figure 5. Percentages of Cues Selected First During Test Session 26

Figure 6. Weighted Pros for Classification Procedure 31

List of tables

Table 1. Relative Frequencies of Cue Values for Friend and Foe Contacts 12

Table 2. Cue Validities of Contacts 13

Table 3. Mean Error and Mean Standard Deviation of the Error of Cue Validity Estimates 17

Table 4. Mean Accuracy, Confidence Ratings, and Response Times 18

Table 5. Mean Accuracy and Confidence Ratings by Pattern Probability 19

Table 6. Conditional Probabilities of Friend/Foe Classification Given Cue Pattern 20

Table 7. Mean Accuracy and Confidence Ratings by Conditional Probability 21

Table 8. Participants' Selection Patterns 23

Table 9. Observed and Predicted Accuracy Levels 32

Acknowledgements

The experiment reported here was completed with the invaluable assistance of Andrea Hawton, Wendy Sullivan, Elaine Macedo, Heather Blunt, and Sarah Young.

Introduction

Background

Building an accurate "picture," or awareness of the situation, is perhaps the most critical aspect of Command and Control (C2). In the context of naval operations, this "situation assessment," involves operators monitoring sensors (e.g., radar, sonar) that provide information about the aircraft and surface and underwater vessels (called "contacts" by operators) in the area around the ship. The operators use this information to detect all craft around them, then classify them in terms of threat level and identify them if possible (i.e. the specific type of craft, its nationality, etc.).

Inferencing is fundamental to threat assessment as it is only by interpreting data and making judgments of the class, position, and intent of entities in the environment that a commander and his/her team can understand the situation. A key concern is the possibility of information overload due to the rapid increase in the amount of sensor data that can be obtained and displayed without a concurrent expansion of the human operator's ability to process information [3]. This problem is confounded by the inherent uncertainty of the battlefield – uncertainty that arises from limitations in information gathering methods and sensors [4] but also the presence of probabilistic processes in the environment [5] [6]. In the latter case, uncertainty reflects not just a mismatch between the information needed by the commander and the information gathering processes employed in the field, but an inherent inability to obtain perfect knowledge of the outcomes that will occur under specific conditions.[1]

To aid naval operators perform the situation assessment process, researchers have been working to develop various forms of decision support systems, such as more informative displays and data analysis tools [e.g., 7]. A constant challenge in this domain, however, is to understand exactly what kinds of information operators need to build more accurate pictures. If a decision support system does not provide information the operator needs and wants or operates in a way that the operator does not understand, there is a danger that the system will not be trusted or used. An even greater danger is that operators who use such a system may misinterpret its outputs and make decisions based on a flawed assessment of the situation.

This research project is intended to help us understand some of the decision making processes used by operators in assessing the situation. Empirical research regarding how people weigh and combine information is not complete, particularly with respect to the kinds of tasks performed in situation assessment. In these experiments, we examine the ways people use different pieces of information (referred to here as "cues") to classify contacts in a simulated naval warfare environment.

Often, information available for building situation awareness is not completely reliable. Sensors can malfunction and are negatively affected by certain environmental conditions that

[1] Of course, even perfect knowledge would not ensure the capability to make perfect inferences, given limitations of human intellect.

cause them to provide inaccurate readings. Also, not all contacts in the same threat class or even of the same type are all exactly alike. Variability in configuration and build means that a characteristic may be associated with a particular kind of contact only some of the time. Thus, a specific radar type or maneuver pattern may be associated with a hostile contact in some but not all cases (e.g., 80% of the time a hostile contact exhibits these characteristics). Moreover, these same characteristics may also be encountered with neutral or even friendly contacts in some cases (say, 5% of friendly craft use the same radar or maneuver pattern as some hostile craft). We call such characteristics "probabilistic cues" because they are associated with a kind of contact at some level of probability or chance. Detecting that cue does not provide unambiguous information about the type of craft or its intent.

Probabilistic cue learning

Numerous studies have examined peoples' abilities to learn how to use probabilistic cues in decision making. These studies have taken many forms but all have been focused on determining how well people can represent the probabilistic relations among cues and a related dimension or criterion through some form of training, then use that knowledge to predict the state of the criterion given subsequent patterns of cue values. The most important factor explored in this line of research has been *cue validity*, which is the extent to which a cue correctly indicates the state of the criterion. Typically, cue validity is defined in terms of the Bayesian probability with which a given state of the criterion should co-occur with a given pattern of cues (the conditional probability of the criterion given the cue pattern). When cues and criterion are continuous variables, researchers often equate cue validity with the regression of the cue on the criterion [8].

A basic phenomenon that has been repeatedly verified is that, given sufficient practice, people can learn the relationships of probabilistic cues to a criterion and make fairly accurate judgments about that criterion on the basis of the cues. Numerous early studies [e.g., 9, 10] have demonstrated that people learn to weight probabilistic cues according to the extent those cues are correlated with a criterion. Often, such studies have examined criteria that have only one or two associated cues but, in some cases, studies have found that people can learn to use many cues and make judgments concerning categorical criteria [e.g., 11]. In cases where cues are linearly related to a criterion, peoples' judgments of the criterion are consistent with a weighted averaging model [12] [13], suggesting that people can internalize some representation of the probabilistic relationships of cues to criterion.

It must be noted, however, that probabilistic cue learning by trial and error is difficult and people are far from perfect in their ultimate performance. For one thing, people use irrelevant cues (i.e., cues unrelated to the criterion) in some situations as well as over-weight highly valid cues but under-weight low validity cues [14] [15]. Indeed, Klayman [16] has noted that many studies indicate limitations on probabilistic cue learning. Among the conditions that strongly impair learning are non-linear relationships of cues to criterion, large numbers of cues, the abstractness of cues, and the nature of feedback provided during learning [e.g., 11, 17]. Another reason peoples' judgments may not exhibit a very high correlation between cue validities and criterion prediction is that people are sensitive to more than just the validity of cues. Newell, Rakow, Weston, and Shanks [18] have found evidence that people combine information about the validity of cues (their probabilistic relation to the criterion) with

information about the discriminability of cues (how frequently a cue distinguishes among alternatives) to make decisions.

Understanding the conditions under which people can effectively learn to use probabilistic cues is important in order to evaluate whether or when peoples' judgments will conform to normative models of probability, and, in turn, how the mind computes these judgments [8]. Recently, debate has risen concerning the interaction of decision making processes with real-world constraints of time, information scarcity, and computational burden to affect the learning and use of cue information. A question of particular relevance to effective C2 is whether decision making strategies are compensatory, weighing all available cues, or non-compensatory. Compensatory and non-compensatory procedures achieve different trade-offs of accuracy and costs in terms of time and cognitive resources. Examining this question will contribute to determination of how command decisions correspond to actual operational constraints of time, information availability, and computational power.

Fast and frugal heuristics

Early, analytic approaches to explaining decision making were based on the premise that human decision making can be modeled in terms of formal processes predicted by normative theories of probability and logic [19]. Numerous sequential and distributed procedures for comparing alternatives are known, most of which can be computationally modeled by production systems operating on a representation of the problem space. Many, for example, are based on Bayesian statistics [20]. A popular general form of analytic theory is the linear compensatory model, which involves the computation of an overall score for each decision alternative based on the sum of relevant dimension values for each alternative, weighted by each dimension's importance [21]. Because the score of each alternative is based on all known dimensions, effects of large and small dimension values can compensate for one another in determining the overall desirability of the alternative [22] [23].

Analytic models, however, often fail to adequately describe peoples' decision making behaviours when performing real-world tasks. Another theoretical approach recently explored in the military context is Naturalistic Decision Making (NDM), a framework that explains decision making in terms of informal recognition-based processes [24]. NDM has become a popular framework for understanding decision making in complex, real-world domains such as military C2 [25]. This approach relies heavily on the notion that decision making critically depends on the quality of the decision maker's situation awareness, or understanding of what is happening around him/her. Situation awareness allows the decision maker to rapidly but accurately match the current situation to past experiences and select a workable course of action.

NDM theories have the advantages of being closely linked to what expert decision makers actually do in real-world domains and being applicable to dynamic, uncertain, and high risk environments, as demonstrated in numerous empirical studies. Their usefulness as models of decision making, however, can be limited by their informal nature, which makes it difficult to develop specific, testable hypotheses for research [26].

One way to deal with the limitations of NDM as a scientific approach is to look to theories of bounded rationality and, in particular, an approach that makes use of "fast and frugal" heuristics [1] [2]. The fast and frugal heuristic approach is based on a conceptualization of rationality in which behaviour is evaluated in terms of its adaptiveness within the limits of time and knowledge imposed by the situation and the computational power and the decision maker [27] [28]. Todd and Gigerenzer [26] define this concept of *ecological rationality* as "adaptive behavior resulting from the fit between the mind's mechanism and the structure of the environment in which it operates."

The basic premise of the fast and frugal heuristic approach is that much of human decision making and reasoning can be explained in terms of simple heuristics that operate within the limits of time, knowledge, and computation imposed on the individual [1]. Thus, this approach combines the explicitness of analytic models with the focus on practical issues of available time and information of NDM [29]. Fast and frugal heuristics do not compute quantitative probabilities or utilities, as in classical decision making models, but are nevertheless implemented as step-by-step procedures consisting of a search rule, stopping rule, and heuristic principles for making the decision [30]. The *search rule* defines the principle by which the heuristic directs its search for alternative choices and for information, whereas the *stopping rule* comprises the principles that specify when and how the search procedure should be terminated. The *heuristic principles for decision making* comprise the procedures used to choose among decision alternatives that have either been presented by the task or generated by the decision maker.

Many different fast and frugal heuristics have been identified as potential solutions to a range of tasks differing in, a) the number of options presented in the decision, b) the number of options that can be chosen, and c) the number and kinds of cues available [1; pp. 29-31]. So-called "ignorance-based decision making" heuristics, for example, are designed for a very simple kind of problem in which the decision maker must select one option from just two possibilities [31]. An example is the Recognition Heuristic, which Todd and Gigerenzer [26] define as follows, "when choosing between two objects (according to some criterion), if one is recognized and the other is not, then select the former." For example, when choosing between two kinds of wine offered by a friend, you might select the vintage that you recognize (either from past experience or reviews of others) as the wine that will taste best. In this case, the sole basis for rejecting the alternative is that you do not recognize it. Although in cases where both or neither object is recognized the heuristic leads only to random choice, the heuristic can yield good choices when only one object is familiar *and* one's familiarity with objects is correlated with their ranking along the judgment criterion [1; pp. 41-43]; i.e. high quality wines actually do receive better reviews and/or are more likely to be sampled at restaurants, etc.

Another example of fast and frugal heuristics is "one-reason decision making," which entails the choice of an option on the exclusive basis of just one cue (or reason) [2]. A one-reason heuristic begins with the selection of a dimension along which to compare options, followed by inspection of the cue values of the two options, and comparison of the options on the bases of those values [1; pp. 77-81]. If the options differ on their cue values, then the process is stopped and the option with greater value is selected. If the options do not differ (or, more realistically, do not differ to a sufficient degree), then the entire procedure is repeated for a new cue dimension until a choice can be made.

The Take-the-Best heuristic

Gigerenzer and Goldstein's [2] "Take the Best" (TTB) heuristic is as a good example of the one-reason decision making. The TTB heuristic works for tasks involving the choice of one of two alternatives based on a single criterion. In their research, for example, Gigerenzer and Goldstein's task was to indicate which of two German cities had the larger population. To make a choice, one would have to either know the populations of the city options or rely on various cues correlated with city population, such as whether the city possessed a professional soccer team. In the latter case, TTB dictates that cues are searched sequentially in the order of their validity or predictiveness until a cue is found that discriminates between the two alternatives. TTB has proven to be a viable strategy for solving the city population problem in simulation studies, performing as accurately, or nearly as accurately, as more computationally intensive linear strategies across all levels of assumed knowledge [2].[2]

The finding that TTB performs comparably in terms of accuracy to linear regression and other compensatory procedures has been replicated with 19 other data sets drawn from psychology, economics, and other fields [1; Ch. 5]. The speed and frugality of TTB derives from its non-compensatory nature [26]. By definition, fast and frugal heuristics employ search rules that limit the number of cues consulted and stopping rules that make choices as soon as sufficient evidence has been obtained. In contrast, most statistical and probabilistic models are compensatory and consult all available cues and make choices only after comparison of multiple options. Fast and frugal heuristics can nevertheless perform accurately because they take advantage of the structure and regularities of information in a particular task environment [1; pp. 113-114]. Thus, TTB performs well when the task environment is structured in a non-compensatory way; i.e. when the validity or importance of cues falls off dramatically in a particular pattern [Gigerenzer et al., 1999, pp. 120-124]. In this environment, the best cue is likely as reliable an indicator of the correct choice as the weighted average of all available cues.

Although research strongly suggests that TTB is an effective decision strategy in certain task environments, there is less evidence that people actually use TTB in performing two-item discrimination tasks. Dhami and Ayton [32], for example, observed that, when making decisions in bail cases, roughly 32% of British magistrates exhibited patterns of decisions consistent with TTB. In another study, Broder [33] asked subjects to classify "alien" creatures into two categories based on sets of probabilistic characteristics, a task like the city population task, except that subjects learned an artificial reference class and cue values. Using a statistical procedure to classify the patterns of classification decisions of individual participants, Broder found that only 28% of subjects' choice behaviors could be classified as consistent with TTB. The remaining participants seemed to use some other strategy that was probably compensatory. In a subsequent experiment, in which Broder required participants to "purchase" cues by expending some amount of resources to uncover cue values, 40% of participants were classified as using TTB when the cost of cue information was relatively low and 60% when the cost was relatively high. This suggests that TTB is a strategy *available* to decision makers but task conditions, especially the costs associated with obtaining

[2] The accuracy of TTB in this task depends on the extent to which cues are predictive of city population. TTB works best when a few highly predictive cues are available.

information, play a large role in determining whether people employ it. It must be noted, however, that Newell and Shanks [34] found less evidence that people use TTB than previous studies, although factors such as the cost of information search, knowledge of the true cue validities, and deterministic associations of cues to criteria affected the proportion of choices consistent with predictions of TTB. Peoples' search for information, in particular, deviated substantially from what would be expected if decision makers were employing a fast and frugal heuristic.

The Take-the-Best-for-Classification heuristic

Threat classification, although frequently difficult and requiring extensive use of sensors and sophisticated sensor-use techniques in actual operations [35], ultimately boils down to a decision of which of a few classes (hostile, potential threat, neutral, friend) a contact should be assigned to. The decision is made on the basis of sensor data that serve as cues to the appropriate classification. The operator must rely on his or her knowledge of how these cues are related to threat class to place the contact into a threat class. Thus, whereas TTB performs the task of choosing between two alternatives on the basis of some dimension, threat classification is a task requiring the placing of a single object into one of two or more threat categories.[3] Nevertheless, some key elements of TTB can be considered for their relevance to threat classification:

1. **Frugal search:** The use of the minimum amount of information needed to reach a decision; and

2. **Simple decision rule:** The rule that is used to assign a contact to a threat class, which must involve little computation.

Threat classification can be performed by a heuristic that adheres to the principle of frugal search. In this case, the minimum information needed is one cue that can indicate the threat class to which a contact likely belongs. The task can also be performed with a simple decision rule of selecting the threat class to which the value of that cue is most strongly associated. The stopping rule for threat classification is built into the decision rule; i.e. search is terminated when a cue is located that can be used to make a decision.

Based on this analysis, a variant of TTB, called Take-the-Best-for-Classification heuristic (TTB-C), was devised to perform the threat classification task.[4] It is based on the premise that the single best cue can be used to make accurate threat classification judgments in a task environment in which that cue is highly predictive. Thus, TTB-C is not intended to be universally applicable but a fast and frugal alternative when one or more cues point to the appropriate threat classification at an acceptable rate.[5] Unlike TTB, which chooses between

---

[3] For the sake of simplicity, the experimental task involved only a binary friend-or-foe classification judgment.

[4] TTB-C is also derivable from the Lexicographic heuristic for two-alternative choice, which is a generalization of Take-the-Best [36; p. 143].

[5] What is an acceptable rate must be determined through the balancing of accuracy demands and resource limitations in terms of time and available information.

two objects along a single dimension, TTB-C places a single object into one of two categories along the threat dimension. Thus, TTB-C is simpler in some respects than TTB but it takes from TTB the basic search concept of locating the single best cue to make its decision. TTB-C is illustrated in Figure 1.

Figure 1. ***The Take-the-Best-for-Classification (TTB-C) Heuristic***

TTB-C, as illustrated here, assumes that there exist one or more cues that have some non-random association to the threat class of contacts and that all, or some subset, of these cues can be inspected by the decision maker. Moreover, the decision maker must have acquired, through experience or training, knowledge of the relative validities of these cues. These, of course, are not minor assumptions but, as the probabilistic cue learning literature has illustrated, there is sufficient reason to believe people can learn cue validities, even if their learning is imperfect.

TTB-C works in the following way. Given an as-yet-unclassified contact, the heuristic begins by searching for the single most valid cue to serve as the basis for classification. In the experiment described in this report, all cues associated with contacts will be available, so the most valid cue should always be inspected. If the experimental procedure made some cues unavailable for certain contacts, the decision maker would have to determine whether the most valid cue was available and, if it was not, then search for the next most valid cue. The decision rule is equally simple; when the most valid available cue is located, the decision maker assesses which threat class has the greater probability of being true given the value of

that cue and makes that threat class the output of the heuristic. The heuristic will be used to make the simplified two-category choice (friend or foe) of the experimental task but could apply to threat classification with the traditional set of threat classes (hostile, potential threat, neutral, and friend). With the contact classified, the heuristic terminates. Should no valid cue be found, the decision maker has only one recourse, which is to guess.

Other fast and frugal heuristics have been proposed to perform categorization. The Categorization by Elimination heuristic is one that applies simple search and stopping rules to select a category designation from multiple options by using successive cues to eliminate more and more alternatives until a single option is left [37]. This heuristic is well-suited to tasks in which an individual must identify the class of an object from an undefined set of potential categories. The threat classification task examined here, however, is highly structured (each contact is represented by four binary cues) and only two threat classes are possible. Thus, TTB-C seems to be faster and more frugal approach to threat classification than even Categorization by Elimination.

Purpose of study

The aims of this experiment were to, specifically, examine the extent to which TTB-C was a viable model for human judgment in a threat classification task and, more generally, assess the extent to which participants' cue selection and classification behaviour would be consistent with the principles of fast and frugal heuristics. Based on the TTB-C heuristic described in the previous section, it is possible to generate several hypotheses concerning peoples' threat classification performance:

1. When provided with trial-and-error learning experience, participants will learn the relative ranking of cues according to their validity in predicting threat class;

2. During a test session in which participants must select a cue to inspect its value, participants will select the most valid cue first for virtually all contacts;

3. Participants will rarely select any additional cues because the most valid cue will always be available and TTB-C can make a decision without additional information (additional cue selections would likely reflect attention lapses and response errors) [38]; and

4. Participants' accuracy rate in judgments of threat class will be roughly 90%, the proportion expected if they exclusively use only the most predictive cue in this experiment.

TTB-C embodies the fast and frugal principles of limited search and simple decision rule. Thus, even if participants do not adhere to the predictions of this particular heuristic, it is of interest to learn whether, and in what ways, their decision making is fast and frugal. A key assumption of TTB-C and other fast and frugal heuristics is that people have reasonably accurate knowledge of the task environment, especially the identity and relative validities of predictive cues. Without such knowledge, there is no basis for these heuristics to order information search. Thus, another important aim of this study was to examine whether subjects are able to accurately learn the underlying cue validities of a stimuli set and use that

information in selecting cue information during a test session. Related to this aim, the impact of the differences between validities of different cues were varied to assess how participants would deal with cue-patterns for which it was relatively easy or difficult to learn the rank ordering of cue validities.

Method

This experiment investigates how people learn to use different pieces of information (cues) to make classification judgments. The experimental task was framed in a simulated threat classification environment. Various "contacts" (simulated craft) were presented on a simulated radar screen for participants to classify as either friend or foe based on the values of four characteristics of the contact (the cues). The availability of just four cues is unrealistic, as was the fact that all the cues were strictly binary, but the task was not intended to accurately describe threat assessment in an actual military context. The experimental task allowed precise variations of the relationships of cues to contact classification. Thus, for this experiment, we created an environment in which objects had to fall into one of two possible classes (friend or foe) and were described by four cues, each of which could take on one of two values. Each cue value had a specific probability of being associated with friend and foe contacts, with these probabilities determining the cue's validity in classifying contacts.

Participants

Participants were 24 men and women who were employees of DRDC Toronto, students conducting research at DRDC Toronto, or individuals recruited from local universities. All received payment in exchange for participation. All participants were aged 18 and older, had normal or corrected-to-normal vision, and were unfamiliar with the specific hypotheses and stimulus configurations of the experiments.

This study, approved by the DRDC Toronto Human Research Ethics Committee, was conducted in conformity with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

Materials

All experiments were conducted with Pentium PC computers, which presented stimuli, collected subject responses, and recorded data. The experimental platform was the Team and Individual Threat Assessment Network (TITAN), which has been used successfully in previous research on individual and team decision making in command and control situations. TITAN was modified somewhat for this study to facilitate the study of probabilistic cue learning.

TITAN is a low fidelity threat assessment simulator. The interface (illustrated in Figure 1) presents a radar screen on which "contacts" are presented by asterisk symbols. Each contact corresponds to a single entity around the participant's "own ship," which is indicated by a blue circle at the center of the radar screen. Using the computer mouse, the participant can click on ("hook") a contact, which activates a set of buttons that allow access to information about that contact. This information consists of four characteristics, such as speed, altitude, and so on. The interface can be customized to allow participants to view all of the contact's characteristics at once or to restrict participants to viewing one characteristic at a time. By clicking another button, participants call up a box in which two possible classifications,

"friend" and "foe," are indicated. Radio buttons under each classification allow the participant to indicate a classification judgment. Further windows open to allow participants to indicate a confidence judgment and receive feedback concerning their classification accuracy.

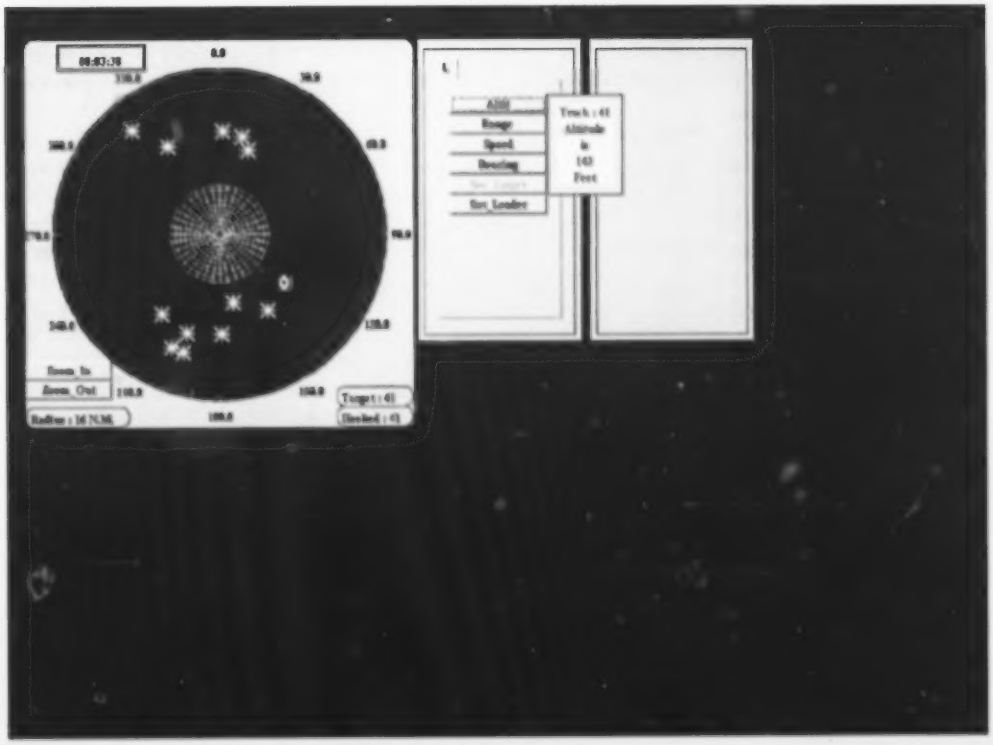

Figure 2. TITAN interface (shows the cue selection screen; separate windows appear for threat classification judgments and confidence rating)

A set of 200 contacts (100 friend and 100 foe) was created for the training session and a set of 100 contacts (50 friend and 50 foe) for the test session. Table 1 indicates the proportions of friend and foe contacts possessing each cue value for the four cues.

Table 1. Relative Frequencies of Cue Values for Friend and Foe Contacts

	HIGH CUE VALIDITY DIFFERENCE CONDITION							
	Cue 1 (Speed)		*Cue 2 (Altitude)*		*Cue 3 (Initial Bearing)*		*Cue 4 (Initial Range)*	
	Value 1 (25-35 kt)	*Value 2 (> 35 kt)*	*Value 1 (0 ft)*	*Value 2 (> 0 ft)*	*Value 1 (91-270°)*	Value 2 (0-90°)	*Value 1 (0-20 nm)*	*Value 2 (> 100 nm)*
Friend	90%	10%	60%	40%	30%	70%	20%	80%
Foe	10%	90%	40%	60%	70%	30%	80%	20%
	LOW CUE VALIDITY DIFFERENCE CONDITION							
	Cue 1 (Initial Climb/Dive)		*Cue 2 (Signal Strength)*		*Cue 3 (Direction of Origin)*		*Cue 4 (Missile Capability)*	
	Value 1 (0)	*Value 2 (> 0)*	*Value 1 (High)*	*Value 2 (Medium)*	*Value 1 (B. Lag.)*	*Value 2 (Red Sea)*	*Value 1 (None)*	*Value 2 (High)*
Friend	82.5%	17.5%	67.5%	32.5%	27.5%	72.5%	22.5%	77.5%
Foe	17.5%	82.5%	32.5%	67.5%	72.5%	27.5%	77.5%	22.5%

Note: B. Lag. = Blue Lagoon

Design

We manipulated two variables in this experiment. The first, varied between subjects, was the Instruction Set given to participants (Describe vs. Discriminate). The Describe instructions indicated that participants were to "learn the characteristics of friend and foe contacts so that you will be able to describe each of these later in the experiment." The Describe instructions were intended to emphasize the patterns of cue values associated with friend and foe contacts. The Discriminate instructions indicated that participants were to "learn how their [friend and foe] different properties, as reflected in cue values, distinguish friend and foe contacts." The Discriminate instructions were intended to emphasize the differences in typical cue patterns between the two types of contact. Although the two instruction conditions differed in the orientation given to participants, the explanation of the TITAN interface and classification task were identical. Annex A contains copies of the two sets of instructions.

The second variable, varied within subjects, was the size of the differences among cue validities (Cue Validity Difference) of the four cues. In one case, the High Cue Validity Difference condition, cue validities differed by increments of 10%, whereas in the Low Cue Validity Difference condition, cue validities differed by increments of 5% (see Table 2). This variable not only manipulated the relative differences of cue validities among cues but also the upper and lower cue validity values of cue sets. The ranges of cue validities, however, centred on 75% in both conditions.

Table 2. Cue Validities of Contacts

CUE VALIDITY DIFFERENCE COND.	CUE VALIDITY			
	Cue 1	*Cue 2*	*Cue 3*	*Cue 4*
High-Difference	90%	60%	70%	80%
Low-Difference	82.5%	67.5%	72.5%	77.5%

Instructions and Cue Validity Difference were counterbalanced by alternating the order in which participants were assigned to the Describe and Discriminate Instruction Set conditions and alternating the order in which each participant completed the High and Low Cue Validity Difference cue sets.

Procedure

Prior to beginning an experiment, subjects received a thorough briefing on the purpose of the experiment, the TITAN software and its use, and the nature of their task.

The experiment was divided into a training and test phase. In the training phase, participants received 200 contacts, of which 100 were friends and 100 foes. All contacts were presented in random positions on the radar screen simultaneously, although the participant was required to use "zoom in" and "zoom out" buttons to view all of the contacts. Each contact had four cues associated with it, specifying cue values generated according to the probability matrix shown in Table 1. Participants selected one contact at a time in any order they chose and accessed that contact's cue values. All four values were available on the screen at the same time but the order in which cues were listed was random from contact to contact. Participants then made a classification judgment, indicating that the contact is either friend or foe. After indicating their classification judgment, participants indicated their confidence in the accuracy of their response on a 100-point scale.[6] After this, participants received accuracy feedback on their classification judgment in the form of a message indicating whether they were correct or incorrect and provision of the correct classification. Participants received no initial information concerning the predictiveness of cues and all learning occurred through trial-and-error.

Following the training phase, participants were allowed a short break then performed the test phase. The test phase followed the same procedure as the training phase with a number of important differences. The key change in procedure was that participants could no longer access all cue information simultaneously. During the test phase, each cue was represented by an individual button that participants pressed to view the value of that cue. The order of the buttons was randomized from contact to contact. Participants were given no specific instructions concerning how many cues to select; they were told to view whatever cue information they wanted before making their classification judgment. In addition, participants were presented with only 100 contacts (50 friends and 50 foes) and they did not receive feedback on the accuracy of their judgments. The classification judgments and confidence

[6] Because the choice was binary, participants required only the upper (50 – 100) range of the scale.

ratings were made in the same fashion as during the training phase. Participants were under no time constraints when making their judgments.

Following the training and test phases, subjects were asked to indicate the diagnosticities of each cue as a predictor of contact classification by judging the proportions of cue values associated with friend and foe contacts.

Results

Participants' performance in the training session and responses to the post-experiment survey were analysed first to determine how well participants learned to classify contacts. Then participants' performance and cue use in the test session were analysed to examine participants' decision and cue selection strategies.

Training session

The contacts presented during the training session were divided into five blocks of 40 contacts each, based on the order of presentation (i.e., the first 40 contacts, the next 40, etc.). Accuracy scores (the percentage of contacts correctly classified as friend or foe) were calculated for each block for each subject to create mean accuracy scores, which are shown broken down by Instruction Set and Cue Validity Difference conditions in Figure 3. A three-way, mixed design (one between-subjects and two within-subjects factors) Analysis of Variance (ANOVA) revealed a significant effect of Training Block [$F(4,88) = 33.75$, $MSe = 71.28$, $p < .05$] but no significant main effects of either Instruction Set [$F(1,22) = 1.59$, $MSe = 634.54$, *n.s.*] or Cue Validity Difference [$F(1,22) = 0.81$, $MSe = 279.15$, *n.s.*]. There were also no significant interaction effects among any of the variables. As can be seen in Figure 3, accuracy generally increased over Trial Block, indicating that participants learned how to classify friends and foes more accurately through trial and error training. Although participants exhibited, for the most part, steady improvement with training there is no clear indication that accuracy scores reached an asymptotic level, which suggests that participants could have improved their performance with further training.

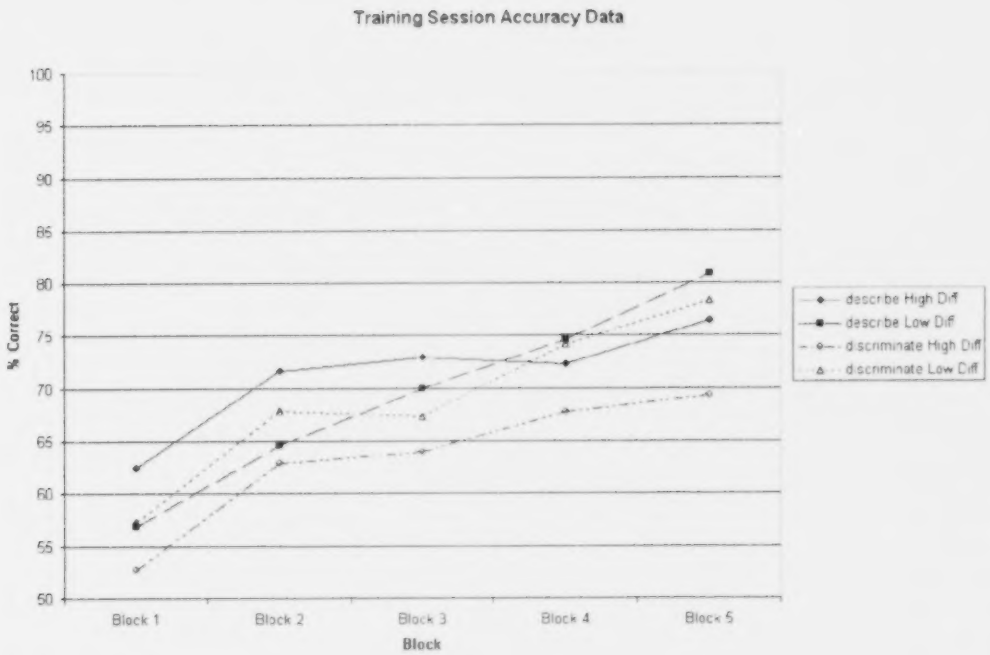

Figure 3. Classification Accuracy by Block in the Training Session

Another measure of participants' degree of learning comes from their responses to the post-experiment survey. Participants were asked to estimate, for each cue, the percentages of friends and foes possessing each of the two possible cue values (the actual percentages are presented in Table 1). Error scores were calculated for each participant by subtracting the actual cue value percentage from the participant's estimate for each cue value and computing the average difference or error score across the four cues. These mean error scores are shown in Table 3 and indicate the average amount by which subjects over- or underestimated the percentages of friends and foes possessing each cue value. In addition, we calculated the standard deviations of participants' error scores, which indicate the variability of participants' error scores around the mean.

***Table 3**. Mean Error and Mean Standard Deviation of the Error of Cue Validity Estimates*

INSTRUCTION SET	CUE VALIDITY DIFFERENCE CONDITION			
	High		*Low*	
	Mean Error (%)	***SD of Error (%)***	***Mean Error (%)***	***SD of Error (%)***
Describe	-1.22	25.30	-0.47	21.40
Discriminate	-2.69	32.77	-1.37	28.75

Note: Negative error scores indicate underestimation of true cue value association probabilities.

A two-factor, mixed-design ANOVA revealed no significant effects on participants mean error scores of either Instruction Set [$F(1,22) = 0.70$, $MSe = 23.84$, *n.s.*], Cue Validity Difference [$F(1,22) = 1.19$, $MSe = 10.85$, *n.s.*], or their interaction [$F(1,22) = 0.09$, $MSe = 10.86$, *n.s.*]. Thus, by this measure, participants in all conditions were equally good at estimating the cue validities of the stimulus set. Moreover, mean error scores are all quite low (post-hoc comparisons revealed no significant differences between mean error scores and zero), suggesting that participants developed highly accurate mental representations of cue validities.

Inspection of the standard deviations of mean error scores, however, reveals that participants' estimates of cue validities were highly variable. A two-factor, mixed-design participant ANOVA revealed a marginally significant effect of Instruction Set on standard deviation of error scores [$F(1,22) = 4.07$, $MSe = 161.73$, $p < .06$] and no significant effects of Cue Validity Difference [$F(1,22) = 2.66$, $MSe = 70.65$, *n.s.*] or the interaction of those two factors [$F(1,22) < .001$, $MSe = 70.65$, *n.s.*]. The marginal effect of Instructions may indicate that participants in the Discriminate Instructions condition, who exhibit the greatest variability, did not internalize cue validities quite as well as those in the Describe instructions condition. Most striking, however, is the size of the means of standard deviations of error scores, which significantly differed from zero for both Describe instructions (collapsed across Cue Validity Differences) and Discriminate instructions. These effects indicate that, although participants' errors averaged out to near zero, their individual judgments of cue validities often exhibited substantial error. This result suggests that participants had learned that the structure of the stimulus set was such that cue validities "balanced" one another (i.e. that for each cue value that was more frequently associated with friend or foe, its complementary value was as frequently associated with the opposite contact class). Participants, however, did not necessarily learn the precise values of cue validities.

Overall, the data from the training sessions and post-experiment surveys indicate that participants had imperfect knowledge of the relationships among cues and contact classifications after completing 200 training items.

Test session

All remaining analyses were performed on data from the test sessions.

Accuracy

Participants' accuracy in threat classification judgments was measured by the percentage of contacts correctly classified. Mean accuracy scores (% correct) are shown in Table 4, broken down by Instruction Set and Cue Validity Difference conditions, as well as the type of contact (Contact Type; i.e. friend, foe, and combined). Although accuracy scores are noticeably larger in the Describe than Discriminate Instruction condition, a three-way, mixed design ANOVA indicated that the main effect of Instruction Set not statistically reliable [$F(1,22) = 2.33$, $MSe = 0.11$, *n.s.*]. Neither was the main effect of Cue Validity Difference reliable [$F(1,22) = 0.54$, $MSe = 0.01$, *n.s.*]. Unexpectedly, the only significant effect revealed by the ANOVA was that of Contact Type [$F(1,22) = 8.72$, $MSe = 0.12$, $p < .05$]. This effect was unexpected because predictability of friend and foe contacts were exactly equal given the balanced construction of cue patterns in the stimulus set; i.e. each cue had a complementary association of cue values to friends and foes. Nevertheless, participants made reliably more accurate judgments for foes than friends. The ANOVA revealed no significant interactions among any variables.

Table 4. *Mean Accuracy, Confidence Ratings, and Response Times*

	CONTACT TYPE								
	Friend			*Foe*			*Combined*		
INSTRUCTIONS	***Acc.***	***Conf.***	***RT***	***Acc.***	***Conf.***	***RT***	***Acc.***	***Conf.***	***RT***
	High Cue Validity Difference								
Describe	.72	79.17	12.57	.80	77.50	11.45	.76	78.34	12.01
Discriminate	66	74.52	14.02	.70	75.24	13.57	.68	74.88	13.79
	Low Cue Validity Difference								
Describe	.74	80.20	12.59	.82	81.64	12.12	.78	80.92	12.36
Discriminate	.69	75.27	12.61	.74	76.43	11.07	.71	75.98	12.29

Acc. = "Accuracy (% Correct)"; Conf. = "Confidence (Rating 0-100)"; RT = "Response Time (sec)"

Because Cue Validity Difference did not affect accuracy, all subsequent analyses were performed on data collapsed across that factor. Although Instruction Set also did not have a significant effect, we have retained this factor in analyses because it was a between-subject variable and aggregating conditions would have increased the sample size and the corresponding probability of a type 2 error (incorrectly failing to reject the null hypothesis).

Given the specific cue validities of contacts, certain combinations or patterns of cue values were more frequent in the stimuli set. The probabilities of each of the 16 possible patterns (given four cues with two possible values) were calculated and divided into high and low probability sets. The high probability patterns consisted of four patterns, each with a greater than 10% chance of occurring for any given contact, whereas the low probability patterns consisted of the remaining 12 patterns, each with a less than 10% chance of occurring for any given contact. High Probability patterns comprised 50% and 49% of the stimuli set for the High and Low Cue Validity Difference conditions, respectively. Low Probability patterns comprised 50% and 51% of the stimuli set for the High and Low Cue Validity Difference conditions, respectively. Thus, participants encountered each individual high probability pattern more frequently during training than any given individual low probability pattern but roughly equal numbers of high and low probability patterns overall.

Table 5 presents participants' mean accuracy scores broken down by Contact Type and Cue Pattern Probability. As can be seen, participants performed more accurately for contacts associated with High Probability (hence more frequent) patterns than Low Probability patterns [$F(1,22) = 170.49$, $MSe = 1.68$, $p < .05$]. The effect of Pattern Probability also significantly interacted with the effect of Contact Type [$F(1,22) = 15.72$, $MSe = 0.06$, $p < .05$]. This interaction effect reflects the somewhat larger difference in accuracy scores between friends and foes for Low than High Probability patterns. Participants were almost as accurate for friends as foes when dealing with High Probability patterns. Nevertheless, participants were more accurate for High than Low Probability patterns overall. Each individual High Probability pattern was more frequently encountered than any given Low Probability patterns, which should have facilitated learning of specific associations of these patterns to their likely contact classifications.

Table 5. *Mean Accuracy and Confidence Ratings by Pattern Probability*

	CONTACT TYPE					
	Friend		***Foe***		***Combined***	
INSTRUCTIONS	***Acc.***	***Conf.***	***Acc.***	***Conf.***	***Acc.***	***Conf.***
	High Pattern Probability					
Describe	.87	81.43	.92	82.13	.90	81.78
Discriminate	.81	75.21	.84	75.82	.82	75.51
	Low Pattern Probability					
Describe	.55	76.39	.70	76.17	.63	76.28
Discriminate	.51	74.82	.62	75.81	.56	75.32

Acc. = "Accuracy (% Correct)", Conf = "Confidence (Rating 0-100)"

Perhaps more important than the frequency with which participants encountered a particular pattern was the degree to which a given cue pattern predicted whether the contact was a friend or foe. Just as the cues varied in their validity, overall patterns of cue values were differentially diagnostic of a contact's class. In particular, certain patterns of cue values strongly indicated either friend or foe, whereas other patterns had lower probabilities of being associated with a given type of contact. Table 6 shows the conditional probabilities of a contact class (friend or foe) given each of the 16 possible patterns of cue values (Table 1 in the Method section provides the names of cues and cue values).

Table 6. *Conditional Probabilities of Friend/Foe Classification Given Cue Pattern*

CUE VALUE PATTERN				**CONDITIONAL PROBABILITIES**			
Cue Values				***High Cue Value Difference***		***Low Cue Value Difference***	
Cue 1	***Cue 2***	***Cue 3***	***Cue 4***	***P(Friend | Pattern)***	***P(Foe | Pattern)***	***P(Friend | Pattern)***	***P(Foe | Pattern)***
1	1	1	1	.41	.59	.48	.52
1	1	1	2	.04	.96	.07	.93
1	1	2	1	.11	.89	.12	.88
1	2	1	1	.61	.39	.80	.20
1	1	2	2	.01	.99	.01	.99
1	2	1	2	.09	.91	.25	.75
1	2	2	1	.22	.78	.36	.64
1	2	2	2	.02	.98	.05	.95
2	1	1	1	.98	.02	.95	.05
2	1	1	2	.78	.22	.64	.36
2	1	2	1	.91	.09	.75	.25
2	2	1	1	.99	.01	.99	.01
2	1	2	2	.39	.61	.20	.80
2	2	1	2	.89	.11	.88	.12
2	2	2	1	.96	.04	.93	.07
2	2	2	2	.59	.41	.52	.48

To examine the impact of the conditional probabilities with which patterns predicted friend and foe, we divided contacts into High and Low Conditional Probability groups (see Table 7). The High Conditional Probability group included those contacts for which the pattern had a greater than 70% probability of predicting the contact class (friend or foe), whereas the Low Conditional Probability group included those contacts for which the pattern had between a 50% to 70% probability of predicting the hypothesis (friend or foe). The conditional probabilities of friend and foe given a specific pattern are, of course, complementary, so a pattern that is highly predictive of friend necessarily indicated a low probability of foe, and vice versa.

Table 7. *Mean Accuracy and Confidence Ratings by Conditional Probability*

	CONTACT TYPE					
	Friend		***Foe***		***Combined***	
INSTRUCTIONS	***Acc.***	***Conf.***	***Acc.***	***Conf.***	***Acc.***	***Conf.***
	High Conditional Probability of Contact Classification Given Cue Pattern					
Describe	.82	79.88	.86	79.72	.84	79.80
Discriminate	.76	74.99	.78	75.80	.77	75.39
	Low Conditional Probability of Contact Classification Given Cue Pattern					
Describe	.47	74.17	.69	73.29	.58	73.73
Discriminate	.44	76.04	.69	75.85	.57	75.94

Acc = "Accuracy (% Correct)"; Conf. = "Confidence (Rating 0-100)"

A three-way ANOVA with Instruction Set as a between-subject factor and Contact Type and Conditional Probability (of contact class given cue pattern) as within-subject factors, revealed a significant main effect of Conditional Probability [$F(1,22) = 36.93$, $MSe = 1.28$, $p < .05$]. Overall, participants made a higher proportion of correct judgments for patterns for which the conditional probability of the contact class was high than for patterns for which the conditional probability was low. There was a significant interaction effect found between Contact Type and Conditional Probability [$F(1,22) = 4.99$, $MSe = 0.24$, $p < .05$]. When patterns had a High Conditional Probability of predicting the contact class, participants were roughly as accurate for judgments of friend and foe. In contrast, when patterns had a Low Conditional Probability of predicting the contact class, participants were much more accurate for judgments of foe than friend.

Although the conditional probabilities associating contact class with certain cue patterns differed between the High and Low Conditional Probability conditions, the expected level of performance predicted by TTB-C was the same. That is, participants should have been able to achieve equal levels of accuracy for both subsets of items by consistently applying TTB-C. This is also true of a strategy based

on the conditional probabilities themselves. Because the task involved assigning a contact to the most likely threat class, a strategy of selecting the class (friend or foe) most probable for a given cue pattern, regardless of the numeric value of the probability, also allows for fairly accurate classification of both subsets of items.

"Contrary" item analysis

Most cue patterns had either a high or low conditional probability of predicting the contact class. The stimulus set, however, contained two special patterns 1, 2, 1, 1 (i.e. Cue 1 possesses Value 1, Cue 2 possesses Value 2, etc.) and 2, 1, 2, 2, which were the only patterns in which the contact class predicted by the pattern as a whole was contrary to the class predicted on the basis of Cue 1 alone (the most predictive cue). Thus, for pattern 1, 2, 1, 1, the conditional probability of the contact being a foe is less than 50%, which leads to a prediction of friend, but the value of Cue 1 is strongly associated (90%) with foe, leading to a prediction of foe. Likewise, for pattern 2, 1, 2, 2, the conditional probability of the contact being a friend is less than 50% (predict foe) but the value of Cue 1 is strongly associated (90%) with friend (predict friend). Although the experimental design had not been developed with these stimuli in mind, they offer the most compelling test of whether participants attempted to integrate all cue values or sought to use only the most predictive cue in making classification judgments. In these cases, participants were forced to choose between the conflicting predictions of the pattern as a whole or Cue 1 alone. Unfortunately, these patterns were rare in the test set and subjects made judgments like this for only a few items. Specifically, a total of 11 items in the High Cue Validity Difference condition allowed the contrast between predictions of Cue 1 alone and the pattern as a whole. For seven of these items, the pattern predicted the correct classification and for the other four items, the value of Cue 1 predicted the correct classification. Only five items in the Low Cue Validity Difference condition allowed the contrast between predictions of Cue 1 alone and the pattern as a whole. The pattern predicted the correct classification for just one of these items and for the other four items, the value of Cue 1 predicted the correct classification. Thus, caution must be taken in interpreting subsequent comparisons because they are based on participants' behaviour on a small number of items.

By examining participants individual choices for the "contrary" stimulus items, we identified whether those choices were in line with the prediction based on the pattern as a whole or the prediction based on Cue 1 alone. We then determined the number of participants who always chose the classification predicted by the pattern as a whole, the number who always chose the classification predicted by Cue 1 alone, and the number who did both for different items. The percentages of participants following each of these strategies are shown in Table 8. As can be seen, only 12.2% overall failed to consistently choose according to the prediction of the pattern or Cue 1. However, among those participants following a consistent strategy, more made their classification judgments according to the pattern as a whole than Cue 1 alone [$F(1,22) = 13.69$, $MSe = 1.93$, $p < .05$]. These data provide evidence that, for the contrary items, a majority of participants attempted to use all available cue information to make classification judgments but a sizeable minority based their decisions on the value of the most diagnostic cue alone.

Table 8. Participants' Selection Patterns

INSTRUCTION SET	CUE VALIDITY DIFFERENCE	% SUBJECTS CONSISTENTLY CHOOSING BY PATTERN	% SUBJECTS CONSISTENTLY CHOOSING BY CUE 1	% SUBJECTS EXHIBITING NO CONSISTENT STRATEGY
Describe	High	66.7%	20.8%	12.5%
	Low	80.5%	8.3%	11.1%
Discriminate	High	50.0%	31.8%	18.2%
	Low	58.3%	33.3%	8.3%
Combined	High	58.7%	26.1%	15.2%
	Low	69.4%	22.2%	8.3%
	TOTAL	**63.4%**	**24.4%**	**12.2%**

Note: Percentages may not sum to 100% due to rounding.

Response time

Response times were measured from the time at which the participant hooked a contact to the time he/she indicated a threat classification and pressed the Return key on the computer keyboard. Thus, response times included time spent inspecting cues as well as time spent deciding on which threat class to assign the contact.

No predictions concerning response times were drawn from the decision strategies under consideration. In fact, with no instructions to participants to minimize their response times for decisions, the current procedure would not be suited to testing any predictions concerning response times. Nevertheless, mean response times were computed for participants and overall means are reported in Table 4. Generally, participants took a fair amount of time, on the order of 12-13 seconds, to inspect cues and indicate their decisions. Thus, participants appear to have approached the task seriously and were diligent in their efforts to classify contacts.

A mixed design ANOVA was performed to determine whether Instruction Set and/or Cue Validity Difference affected the speed with which participants performed classifications. The only significant effect revealed was, unexpectedly, of Contact Type [$F(1,22) = 7.98$, $MSe = 19.14$, $p < .05$], but no significant effects of Instruction Set [$F(1,22) = 0.33$, $MSe = 9.58$, *n.s.*] or Cue Validity Difference [$F(1,22) = 0.84$, $MSe = 15.54$, *n.s.*] were observed. In addition, the ANOVA revealed no significant interaction effects among any variables. Thus, participants responded somewhat faster overall for foes than friends, although none of the decision strategies considered would predict any difference in the steps needed to classify a contact as either friend or foe. It is unclear what, if any, importance might be attached to this finding.

Confidence ratings

Mean confidence ratings broken down by Instruction Set, Cue Validity Difference, and Contact Type are shown in Table 4. A three-way, mixed design ANOVA

revealed no significant main effects or interaction effects for confidence ratings. A subsequent ANOVA found that Pattern Probability (see Table 5) significantly affected confidence ratings [$F(1,22) = 9.97$, $MSe = 194.3$, $p < .05$], as participants gave higher ratings to High Probability cue pattern items. This effect is in line with participants' greater accuracy for these items and indicates that participants were, to some degree, aware of their better performance. Pattern Probability significantly interacted with Instruction set [$F(1,22) = 8.64$, $MSe = 168.5$, $p < .05$]. Closer inspection of participants' ratings revealed that, whereas confidence ratings were larger for High than Low Probability Patterns in the Describe instruction condition, there was no significant difference in the Discriminate condition. Participants were actually more accurate for High Probability patterns in the Discriminate condition, so this interaction suggests that, for some reason, participants were less able to correctly assess their level of performance when given Discriminate instructions. Pattern Probability did not interact with any other factor.

Conditional Probability (see Table 7) had a significant main effect on confidence ratings [$F(1,22) = 6.04$, $MSe = 182.9$, $p < .05$] and interacted with Instruction Set [$F(1,22) = 8.69$, $MSe = 263.1$, $p < .05$] but no other factors. Participants gave higher confidence ratings to contacts for which the conditional probability of the classification given the cue pattern was High than Low in the Describe instruction condition but not in the Discriminate instruction condition. Again, participants in the Discriminate instruction condition appear to have been less able to correctly assess their level of performance than those in the Describe instruction condition.

Cue selections

When a participant clicked on a cue button in the test session to inspect the value of that cue, that action was recorded as a "cue selection." All such cue selections were recorded for every test item to determine which cues participants inspected and the order in which they were inspected. One hypothesis examined in this experiment was that participants, having learned the relative validities of cues, would inspect the most valid first. Moreover, it was expected that participants would rarely, if ever, inspect other cues. Figure 4 shows the average number of cue selections by the selection order (1^{st} through 4^{th} cue selected). As can be seen, participants selected at least one cue (1^{st} selection) for virtually 100% of test items (but there were several instances of a participant making a judgment without inspecting any cue). Participants also selected a second, third, and fourth cue for a majority of contacts, although the average number decreases significantly from the first to fourth selections [$F(3,66) = 20.32$, $MSe = 8351$, $p < .05$].

The overall number of cues selected was affected by Cue Validity Difference [$F(1,22) = 10.48$, $MSe = 4006$, $p < .05$], with participants selecting more cues overall in the Low Cue Validity condition, as is evident in Figure 2. Moreover, Cue Validity Difference interacted significantly with the order of cue selections [$F(3,66) = 3.73$, $MSe = 751$, $p < .05$], which can be seen in Figure 4 in the contrast between the relatively large drop in number of selections after the second selection in the High Cue Validity Difference condition and the smaller decline evident in the Low Cue Validity Difference condition. Although the sequence of cue selections did not interact significantly with Instruction Set alone, the three-way interaction of

factors was significant [$F(1,22) = 8.69$, $MSe = 263.1$, $p < .05$]. This three-way interaction seems to reflect a difference in the rate at which number of selections decreases with selection sequence between Describe and Discriminate Instruction conditions in the High but not Low Cue Validity Difference condition.

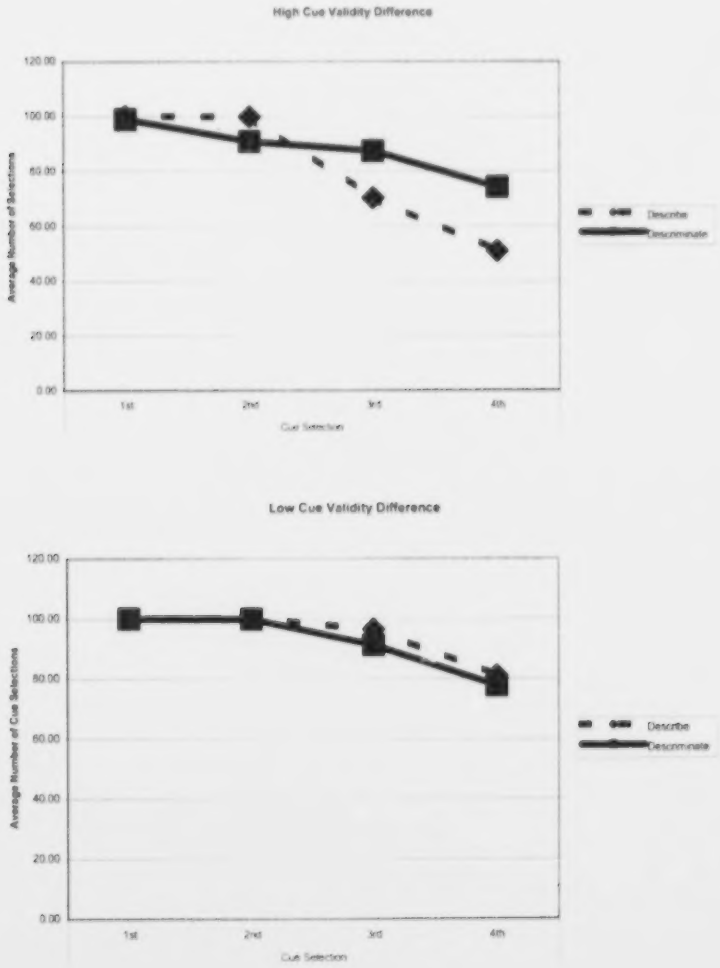

Figure 4. Average Number of Cues Selected by Sequence

Although participants did not search the cues in a frugal manner (i.e. by selecting only the single best cue), they may have sought the most valid cue (Cue 1) first but then inspected others to seek confirmation or to build confidence in their decision (see Newell, Weston, & Shanks, in press). The data, however, do not bear out this possibility. Figure 5 shows the mean percentage of times each of the four cues was selected first by participants. Although participants did often select Cue 1 first, they did not exhibit any clear preference for that cue. Indeed, in both the Describe and Discriminate instruction conditions and for both High and Low Cue Validity Differences, participants more frequently selected a less valid cue first.

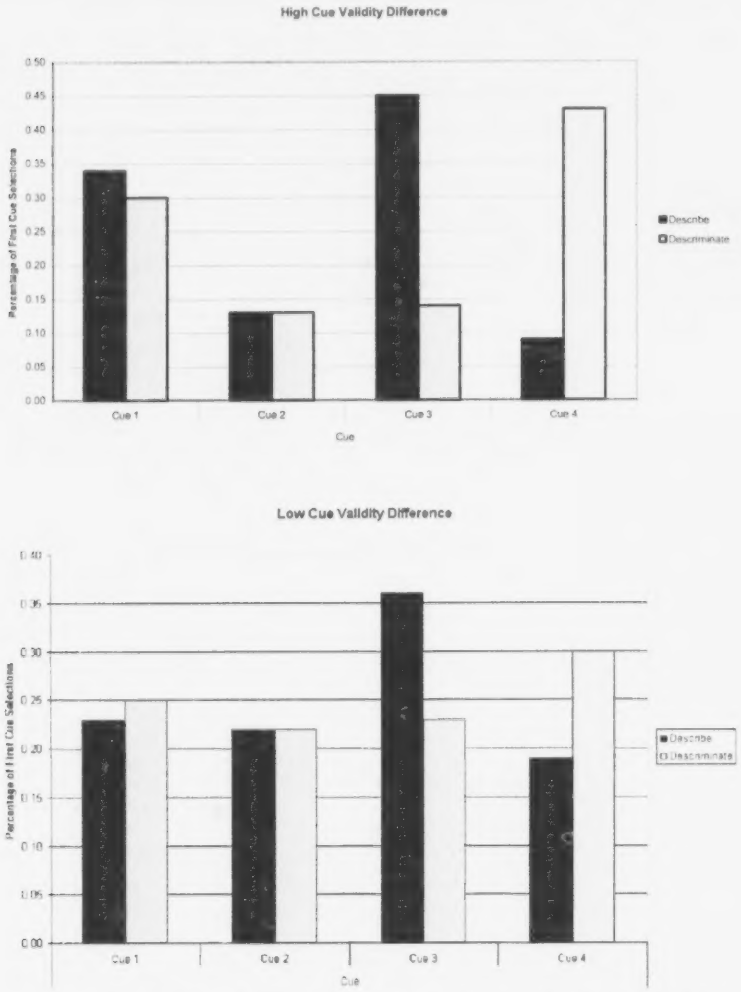

Figure 5. *Percentages of Cues Selected First During Test Session*

A three-way ANOVA with Instruction Set as a between-subject factor and Cue (1-4) and Cue Validity Difference as within-subject factors revealed a significant interaction of all factors [$F(3,66) = 5.00$, $MSe = 0.28$, $p < .05$]. Participants made Cues 1 through 4 their first selection roughly equal numbers of times in the Low Cue Validity Difference condition, which suggests they selected randomly (consistent with the random ordering of cue buttons on the interface screen). In the High Cue Validity Difference condition, participants exhibited a more complex pattern. They often made Cue 1 their first selection for both Instruction Sets and, likewise, selected Cue 2 (least predictive) first relatively infrequently, suggesting that participants did apply some non-random selection strategy. The frequency with which participants selected Cues 3 and 4 first, however, varied considerably between Instruction Sets; Cue 3 was selected first most often in the Describe condition but Cue 4 most frequently in the Discriminate condition. It is not immediately clear what, if any, significance this pattern may indicate. It is unlikely that the effects of Cue Validity Difference reflect interference between the two training trials performed by participants as the High and Low Cue Validity Difference conditions employed separate cue sets and participants performed the second condition following a lunch break.

Even at an individual level, participants exhibited little tendency to systematically select Cue 1 first. Only four participants in the High Cue Validity Difference condition and one participant in the Low Cue Validity Difference condition selected Cue 1 first more than 80% of the time. The other participants selected Cue 1 first at a rate close to chance (25%). One participant, however, did exhibit undeniable evidence of a TTB-C strategy, but only in the High Cue Validity Difference condition. This participant not only selected the most valid cue first on 99 of the 100 test items but also selected a second cue only 8 times and never selected more than that. Thus, this participant could only have used the best cue on the majority of items. This one participant, however, exhibited no such pattern in the Low Cue Validity condition, and no other participant came close to the same level of consistency in favouring the most valid cue.

Discussion

Evaluation of TTB-C

This experiment tested four specific predictions of the TTB-C heuristic concerning participants' performance in a simulated threat classification task:

1. When provided with trial-and-error learning experience, participants will learn the relative ranking of cues according to their validity in predicting threat class;

2. During a test session in which participants must select a cue to inspect its value, participants will select the most valid cue first for virtually every contact;

3. Participants will rarely select any additional cues because the most valid cue will always be available and TTB-C can make a decision without additional information (additional cue selections would likely reflect attention lapses and response errors) [38]; and

4. Participants' accuracy rate in judgments of threat class will be roughly 90%, the proportion expected if they exclusively use only the most predictive cue in this experiment.

Overall, the data did not provide much support for any hypothesis other than the first. Participants did seem to learn the relative validities of the four cues, as evidenced by their estimates of the probabilistic relations of each cue to threat class. Their estimates, however, exhibited a large degree of variability, which indicates each participant did not necessarily learn a completely accurate ranking of cue validities. Participants may have focused more on patterns of cue values than the predictiveness of each cue individually. This would be consistent with their cue selection data and possibly explain why participants' estimates of individual cue validities showed so much variability.

The remaining three hypotheses were clearly falsified. Participants' average accuracy rates were well below the 90% level they could have achieved by following TTB-C (Hypothesis 4). Examination of participants' cue selection patterns revealed that, first, participants generally did not select the most valid cue first more often than would be expected by chance (Hypothesis 2), and second, participants selected more than one cue (and often all four cues) for inspection more often than would be expected due to error or attention lapses (Hypothesis 3). Thus, participants were neither frugal nor systematic in their cue selection, aside from one participant who's behaviour did conform to these two hypotheses in one condition.

Why did participants' behaviour not correspond to these predictions? If participants had not accurately learned the relative cue validities, their low accuracy could be explained as a consequence of TTB-C search being conducted in an inappropriate order. This explanation, however, does not seem to be adequate. Participants' learning was clearly imperfect, as indicated by the large variability of participants' cue validity estimates, but they nevertheless were generally able to indicate a correct ordering of cue validities on the post-experiment

survey. Moreover, participants exhibited no systematic error in their estimates of the true probabilities with which cue values were associated with friend and foe contacts. Thus, it is unlikely that the majority of participants did not realize that Cue 1 was highly predictive of contact class. Yet, this knowledge apparently was not enough to motivate participants to employ TTB-C.

The cue selection data is, perhaps, easier to understand. Participants were under no time pressure in the test session and there were no costs associated with inspecting cues (which is reflected by the frequency with which it was found that participants revisited cues again and again for a substantial number of items). Thus, subjects may have felt free to inspect more cues than they actually needed or used in making their classification judgments. Although this could potentially have hidden use of TTB-C, there is no evidence that participants actually often used that strategy because participants showed no trend toward selecting the most valid cue first nor does their accuracy suggest this either. The best evidence for use of TTB-C comes from the analysis of contrary items, which suggests that some participants, albeit a minority, favoured the most valid cue over the entire pattern of cue values.

In light of the overall failure to find evidence that participants employed TTB-C as a strategy for threat classification, one might question why anyone would use this strategy for this task. There are at least two reasons, based on arguments of Gigerenzer et al. [1], why TTB-C is a plausible strategy for this task. First, TTB-C offers the decision maker great cognitive economy by requiring that only the relative cue validities be learned (as opposed to exact conditional probabilities) and relatively little computation to make a classification judgment. Having learned that Cue 1 was more valid than other cues, participants could have simply classified contacts by the value of that cue and performed the task quickly and with little effort. A second rationale for TTB-C is that a fairly high level of performance was possible despite its economy. The next section features a discussion of other decision strategies and compares the levels of accuracy expected by using each strategy to classify the test stimulus set used in the experiment. TTB-C actually provides a higher level of accuracy in the High Cue Validity Difference condition than that predicted by a strategy of using Bayesian probabilities to make judgments and a comparable level of accuracy in the Low Cue Validity Difference condition. Of course, it only makes sense to use TTB-C if participants *believed* the validity of Cue 1 allowed for an adequate level of accuracy, which appears not to be the case. No instructions were given to suggest that participants should be satisfied with a 90% accuracy level, although participants were warned that 100% accuracy was practically impossible. Nevertheless, participants may have believed they could achieve better performance through some Bayesian or compensatory process.

If there are good reasons to employ TTB-C, why did participants not use cue information in a fast and frugal manner? As noted, they may not have believed that such an heuristic would lead to an adequate level of performance, either because they had not acquired a sufficient representation of cue validities or they believed the experimental task required a greater level of accuracy than even the most valid cue alone could provide. In this latter case, a sense of social obligation may have driven participants to attempt more complex calculations even if they recognized the potential of a fast and frugal heuristic to perform the task. It is possible that participants believed that weighing all cue values would produce better performance in all cases. If participants felt no such obligation they may have underestimated the relative complexity of a weighted averaging strategy in relation to TTB-C, believing that the former

would not be significantly more difficult than the latter. Similarly, participants may have overestimated their ability to weight cues and believed they would achieve a much higher level of accuracy than they actually did. In future experiments, it will be important to survey participants concerning their beliefs about their knowledge of the stimuli set and the relative effectiveness of various classification strategies.

Other decision rules

Just as TTB-C is an adaptation of the TTB heuristic to the single-choice classification problem, other two-alternative choice decision strategies can be adapted. Among the decision strategies that have been examined are Franklin's Rule, Dawes' Rule, and Weighted Pros [36; p. 143]. These are all compensatory procedures for choosing between two alternatives on the basis of cue values, and all have been examined as models for human choice in that type of task. Although not necessarily fast and frugal, the procedures serve as plausible models of choice. Franklin's rule is a procedure by which a decision maker calculates the sum of cue values weighted by the corresponding cue validities for each alternative and selects the alternative with the highest score. Dawes' rule is similar and calculates the sum of un-weighted cue values and selects the alternative with the highest score. The weighted pros procedure examines each cue value for each alternative to determine whether it is consistent or supportive of the alternative (i.e. it is a "pro" for that alternative) and calculates the number of pro cues for each alternative and selects the one with the greater number, guessing if the alternatives have equal numbers of pros.

Versions of these decision strategies can be formulated for the threat classification task. These adapted strategies, unlike their progenitors, do not compare cue values for two alternatives but rather compute a sum of cue values as evidence toward a friend or foe classification and use that value to place the contact in the friend or foe category, depending on the associations of cue values to threat class. Figure 6 contains an illustration of the Weighted Pros for Classification procedure developed for threat classification. The procedure is compensatory, employing all available cues, and does not require a specific search strategy (i.e. there is no reason to predict that cues will be inspected in the order of their validities). Weighted Pros for Classification begins by looking for a relevant cue. If no cue is available, the procedure is forced to guess but if a cue is found, its value is multiplied by its cue validity to produced a weighted value. In the threat classification task investigated here, the cues are binary and their values can be represented as +1 and –1; i.e. either pointing to a classification of friend or of foe. Thus, the next steps in the procedure are to assess whether the weighted cue value is pro for category A (i.e. friend) and/or category B (i.e. foe). If the cue is pro for a category, an evidence sum (assumed to be zero to begin) is incremented by the weighted cue value. When a cue value is associated with both classifications, both friend and foe classes can be supported. Thus, a given cue can lead to increments in evidence for both friend and foe. The next step is to look for another relevant cue. If one is available, the weighting and incrementing steps are performed again. If not, the evidence sums for the two categories are compared to see which is larger. The procedure assigns the classification corresponding to the larger evidence sum or, if the sums are equal, guesses.

The Weighted Pros for Classification procedure just described turns out to be equivalent to the classification procedure one would derive by adapting the Franklin's rule procedure for two-

alternative choice. A classification version of Dawes' rule is performed just as illustrated in Figure 6 but without the weighting step following the selection of a cue.

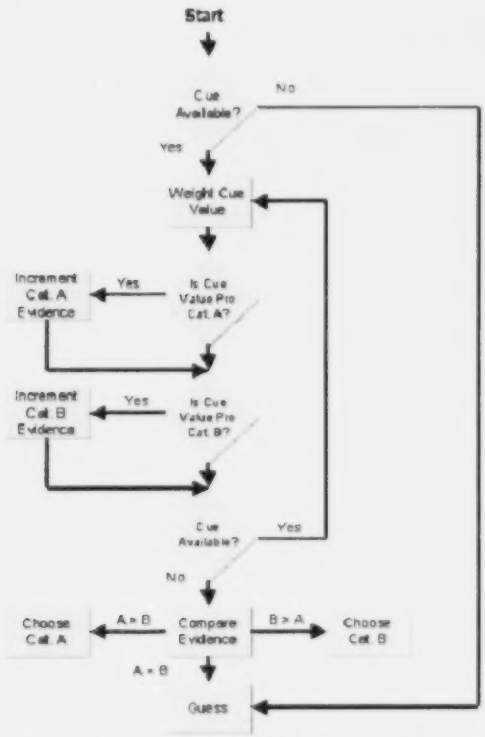

Figure 6. *Weighted Pros for Classification Procedure*

The expected levels of performance for the actual test stimuli used in the threat classification task (test session) were computed on the bases of TTB-C and these other adapted procedures. Table 9 shows these predictions in contrast to the observed levels of accuracy in each condition, as well as predictions based on the Bayesian conditional probabilities of contact classification given the associated cue pattern; i.e. a strategy of choosing the threat class that has the higher conditional probability of being true given the pattern of available cue values. As can be seen, all procedures predict similar levels of accuracy, although a modified Dawes' rule, which does not weight cue values, generally predicts the lowest level of accuracy in the High Cue Validity Difference condition. TTB-C and the modified Dawes' rule make almost identical predictions for the Low Cue Validity Difference condition. Participants' observed accuracies were lower than predictions in all conditions, which may indicate that they had not had enough training to accurately learn cue validities.

Table 9. Observed and Predicted Accuracy Levels

Condition	Observed Accuracy	Predicted Accuracy			
		TTB-C	Weighted Pros	Dawes' Rule	Bayesian Prob.
High Cue Validity Difference					
Describe Instructions:					
Friend	.72	.90	.85	.81	.84
Foe	.80	.90	.92	.85	.90
Combined	.76	.90	.88	.83	.87
Discriminate Instructions:					
Friend	.68	.90	.85	.81	.84
Foe	.70	.90	.92	.85	.90
Combined	.68	.90	.88	.83	.87
Low Cue Validity Difference					
Describe Instructions:					
Friend	.74	.83	.86	.82	.86
Foe	.82	.82	.88	.87	.86
Combined	.78	.84	.87	.84	.86
Discriminate Instructions:					
Friend	.69	.83	.86	.82	.86
Foe	.74	.82	.88	.87	.86
Combined	.71	.84	.87	.84	.86

One cannot distinguish the various decision strategies given the information structure underlying the stimuli used in this experiment. Nevertheless, computing predictions of these strategies will be valuable in further experiments in which stimuli and test conditions are designed to specifically elicit very different predictions of accuracy in classification judgments. Thus, it will be possible to use accuracy data as well as participants' information search patterns to study threat classification judgments.

The relevance of fast and frugal heuristics to issues of Command and Control

Supporting situation awareness will be a priority in C2 for all branches of the Canadian Forces [39]. So far, a great deal of attention has focused on expanding the capabilities of sensors, so that more kinds of information can be gathered with greater precision and across greater ranges, and the sophistication of data processing to enhance the precision and usefulness of sensor data. An unintended consequence, however, has been the rapid expansion of the amount of information the human operator must deal with; an expansion that has not been met by any change in human information processing capabilities [3]. To address this growing tension between what information can be gathered and what the human decision maker can do with it, it is important to understand the kinds of cognitive processes the human mind brings with it to C2 tasks and how those processes can be best employed (or enhanced through training) within the information structure of the warfare domain.

Warfare is complex but necessarily occurs in an environment constrained by universal physical laws, not-quite-so-absolute principles of warfare, and various non-random cultural, political, and historical contexts. Commanders have always confronted uncertainty in these areas but gathering more and more data alone is not way to resolve that uncertainty. In the concept of ecological rationality, Gigerenzer and colleagues [1] show us that uncertainty is, perhaps, best resolved through the match between the structure of one's decision processes and the structure of information inherent in the environment. That is, C2 procedures must be based not on a scatter-gun concept of ever-increasing data collection but on an approach of capitalizing on the structure of the environment and developing faster and more frugal means assess the environment and select appropriate courses of action.

Uncertainty arises not just from a lack of information but from inherently probabilistic processes in the environment. For this reason, it is important to understand the kinds of mental processes people bring to bear on tasks such as threat classification. Threat classification is one of the many tasks that make up situation assessment activities and accurate situation awareness depends on accurate classification judgments. Timely or up-to-date situation awareness, however, depends on classification judgments being made quickly. Where the environment is structured, fast and frugal heuristics can effectively balance these competing aims.

The experiment reported here is the first step in a program to explore how people use probabilistic information to perform threat classification. Recent developments in theories of human decision making offer opportunities to describe, in detail, the cognitive processes by which sensor operators search for and evaluate data when classifying entities. Although this experiment is only a beginning, it is important to continue this line of research to identify the limits of human information processing capacities, cognitive predispositions to either heuristic or analytic reasoning, and concepts for best supporting natural decision making.

References

1. Gigerenzer, G., Todd, P. M. & The ABC Research Group. (1999). Simple heuristics that make us smart. Oxford, UK: Oxford University Press.

2. Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

3. Wallace, J. (2000). The ghost of Jomini: The effects of digitisation on commanders and their working headquarters. In M. Evans & A. Ryan (Eds.), The human face of warfare. Crow's Nest, Australia: Allen and Unwin.

4. Gulick, R M., & Martin, A. W. (1988). Managing uncertainty in intelligence data – An intelligence imperative. In S. E. Johnson and A. H. Levis (Eds.), *Science of command and control: Coping with uncertainty*, pp. 10-18. Washington, DC: AFCEA International Press.

5. Levis, A. H., & Athans, M. (1988). The quest for a C^3 theory: Dreams and realities. In S. E. Johnson & A. H. Levis (Eds.), *Science of command and control: Coping with uncertainty*, pp. 4-9. Washington, DC: AFCEA International Press.

6. Schmitt, J. F., & Klein, G. (1998). Fighting in the fog of war: Dealing with battlefield uncertainty. *Human Performance in Extreme Environments, 3*, 57-63.

7. Paradis, S., Breton, R., & Roy, J. M. J. (1999). Data fusion in support of dynamic human decision making. *The 2nd International Conference on Information Fusion – Fusion '99*. Sunnyvale, CA, USA.

8. Stevenson, M. K., Busemeyer, J. R., & Naylor, J. C. (1990). Judgment and decision-making theory. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology (2nd Ed.), Volume 1*, pp. 283-374. Palo Alto, CA: Consulting Psychologists Press.

9. Brehmer, B. (1973). Single-cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion. *Organizational Behavior and Human Decision Processes, 9*, 377-395.

10. Himmelfarb, S. (1970). Effects of cue validity differences in weighting information. *Journal of Mathematical Psychology, 7*, 531-539.

11. Klayman, J. (1988). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 317-330.

12. Brehmer, B. (1973). Effects of cue validity on interpersonal learning of inference tasks with linear and nonlinear cues. *American Journal of Psychology, 86*, 29-48.

13. Castellan, N. J. Jr. (1992). Relations between linear models: Implications for the Lens Model. *Organizational Behavior and Human Decision Processes, 51*, 364-381.

14. Brehmer, B., & Johansson, R. (1977). Subjects' ability to use information about cue validity in multiple-cue judgment tasks. (Umea Psychological Reports No. 118). University of Umea.

15. Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1083-1119.

16. Klayman, J. (1984). Learning from feedback in probabilistic environments. *Acta Psychologica, 56*, 81-92.

17. Lindberg, L-A., & Brehmer, B. (1977). Effects of task information and active feedback control in inductive inference. (Umea Psychological Reports No. 123). University of Umea, Umea, Sweden.

18. Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2002). Search strategies in decision-making: The success of 'success.' Unpublished manuscript. Department of Psychology, University College London.

19. Gigerenzer, G. (1997). Bounded rationality: Models of fast and frugal inference. *Swiss Journal of Economics and Statistics, 133*, 201-218.

20. Klein, G., (1992). *Decision making in complex military environments.* (Final Contract Summary Report, prepared for Naval Command, Control and Ocean Surveillance Center). Fairborn, OH: Klein Associates.

21. Hunt, E. (1994). Problem solving. In R. J. Sternberg (Ed.), *Thinking and problem solving*, pp. 215-232. San Diego, CA: Academic Press.

22. Ben-Zur, H. (1998). Dimensions and patterns in decision-making models and the controlled/automatic distinction in human information processing. *European Journal of Cognitive Psychology, 10*, 171-189.

23. Hogarth, R. M. (1980). Judgment and choice: The psychology of decision. New York: John Wiley.

24. Klein, G. (1997). An overview of naturalistic decision making applications. In G. Klein & C. E. Zsambok (Eds.), *Naturalistic decision making*, pp. 49-59. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

25. Leedom, D. K., Adelman, L., & Murphy, J. (1998). Critical indicators in Naturalistic Decision Making. *Fourth conference on Naturalistic Decision Making*. Warrenton, VA: Klein Associates.

26. Todd, P. M., & Gigerenzer, G. (2001). Precis of *Simple heuristics that make us smart*. *Behavioral and Brain Sciences, 23*, 727-780.

27. Tietz, R. (1992). Semi-normative theories based on bounded rationality. *Journal of Economic Psychology, 13*, 297-314.

28. Todd, P. M. (2001). Fast and frugal heuristics for environmentally bound minds. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox*, pp. 51-70. Cambridge, MA: The MIT Press.

29. Bryant, D. J. (2002). Making naturalistic decision making "fast and frugal". *Proceedings of the 7th International Command and Control Research and Technology Symposium*. Quebec City, PQ: Command and Control Research Program, Department of Defense.

30. Gigerenzer, G. & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox*, pp. 1-12. Cambridge, MA: The MIT Press.

31. Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*, 75-90.

32. Dhami, M. K. & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14*, 141-168.

33. Broder, A. (2000). Assessing the empirical validity of the "Take the Best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1332-1346.

34. Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 53-65.

35. CMC Electronics Inc. (2002). *Task analysis of the HALIFAX class Sensor Weapons Controller (SWC) and Assistant Sensor Weapons Controller (ASWC) positions: Mission, function and task analysis report*. (DCIEM CR 2002-024). Defence and Civil Institute of Environmental Medicine, Toronto, ON, Canada.

36. Reiskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart*, pp. 141-167. Oxford, UK: Oxford University Press.

37. Berretty, P. M., Todd, P. M., & Blythe, P. W. (1997). Categorization by elimination: A fast and frugal approach to categorization. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, pp. 43-48. Mahwah, NJ: Lawrence Erlbaum Associates.

38. Newell, B. R., Weston, N. J., & Shanks, D. R. (in press). Empirical tests of a fast and frugal heuristic: Not everyone "takes-the-best." *Organizational Behavior and Human Decision Processes*.

39. Department of National Defence (1999). *Shaping the Future of the Canadian Forces: A Strategy for 2020.* Department of National Defence, Canada.

Annex A

Learning Probabilistic Cues Instructions (Discriminate Condition)

Preamble

This experiment investigates how people learn to use different pieces of information (cues) to make judgments about what an object is. In particular, we want to learn how people deal with cues that are not completely reliable. We are using a naval identification and threat assessment computer task to simulate a defence mission aboard a Navy ship. Before we begin the training session, I will describe the task and provide a demonstration of how to operate the simulator.

Instructions

This task may seem complicated at first but we will describe how it works in detail and allow you to practice before beginning the experiment. Please stop me at any time if you have a question.

Let's begin with an overview of the task:

> You are playing the role of an operator in the Operations Room of a naval ship, which is displayed as a blue symbol in the center of your radarscope. Surrounding your ship are asterisks called "contacts." These contacts represent traffic detected by your ship's sensors. Your job is to clear all contacts in your ship's vicinity by assessing their threat levels. Until you classify the contact as friend or foe, the threat they pose is unknown. Threat, in this context, means whether the contact is an enemy who could attack the ship. Each contact's threat classification is based on several information items that you have access to on your workstation. The information items are called "cues" and come from various sensors on the ship. You will have 4 cues to use in making your judgment whether a contact is friend or foe. Friend and foe contacts are different kinds of craft, with different properties that are reflected in different cue values from the sensors. To make your judgments you will first select a contact for classification, review its cues, then indicate whether it is friend or foe. For you to make accurate judgments, you will have to learn how their different properties, as reflected in cue values, distinguish friend and foe contacts. The first session will be a training session in which you will learn by trial-and-error how the cues predict the threat classification of objects. It is important to understand that the cues will never be completely reliable – that is, due to occasional errors in the sensors and variations in the exact configurations of the contacts, no particular cue will always be

100% associated with either a friend or foe contact. After making your judgment for a contact, you will rate how confident you are that your judgment is correct. This procedure will be repeated for each contact.

Some of the information items (cues) will refer to continuous variables, such as speed or altitude, which can take on a wide range of specific values. In this task, however, the exact value of a continuous variable is not important. All continuous variables have a dividing point that cuts the range of values in two. This makes variables like speed binary – i.e. greater than or less than, higher or lower, above or below, etc. – and all you need to learn is which category is associated with friend and which is associated with foe. For example, a speed greater than 300 nm/hr might be a feature often seen for a foe, whereas a friend might be more likely to have a speed less than 300 nm/hr (this is just an example; if speed is a cue in your experiment, don't expect it to behave the way I just said). We have prepared a sheet to indicate what the dividing point is for all continuous variables to help you in this task.

The purpose of the training session is strictly for you to learn how to identify friend and foe. We will not be looking at your performance in the training session, so you need not worry about how many contacts you get correct. It is important to bear in mind that, because no cue is 100% associated with either friend or foe, you cannot expect to ever achieve perfect performance. That is, it is not necessarily possible for anyone to be correct all the time. We just want you to learn to judge friend and foe as accurately as possible.

[Be sure that TITAN is loaded and ready to start]

Demonstration

I will now show you the main features of the computer interface and demonstrate how the task is performed.

Main Features

Feature	*Description*
Pointer	The arrow on the screen is a pointer that allows you to view menu items. You will use the mouse to move the pointer around your screen.
Radar Scope	The black circle within the large gray box is the "radar scope," which displays your ship and the contacts within its vicinity.
Ownship	Your ship is located in the center of the shaded circle in the radar scope.
Contact	Surrounding your ship are asterisks called "contacts." A contact is an object detected by your ship's sensors that appears on your radar scope.
Zoom In & Zoom Out	These buttons allow you to magnify and minimize the range of the radar scope. You can zoom in as close as 1 nm and zoom out as far as 1024 nm. These buttons are used to bring an out-of-range contact into view.
Menus	Next to the radar scope are buttons that permit you to inspect cues for the selected contact.
Information Items	These items are the cues for the selected contact and provide information about the contact. **The cues will always appear in a random order for each contact.**
Set Contact	This button allows you to view all the cues for the selected contact at the same time.
Set Threat	This button is used when you are ready to submit your threat assessment. It calls up the Classification Menu.
Classification Menu	This menu is used to indicate the selected contact's threat classification. You make your decision by clicking the appropriate button under the Friend and foe labels.
Confidence Bar	This bar is used to indicate how confident you feel in your classification judgment, on a scale of 0 to 100. You indicate your rating by clicking on the bar and sliding the pointer up or down until the rating shown at the side is the rating you wish to give.
Short and Cumulative	The short average refers to the average error rate of

Averages	the 6 most recent trials. The cumulative average refers to the average error rate across all trials. These averages are provided solely to help you in learning to classify friend and foe.

Step-by-Step Instructions

There are several steps involved in clearing a contact from the display:

1. **Selecting a contact:** First, you must choose a contact. There are no rules governing how you make your selection. You can select randomly, or start at the most zoomed in setting, selecting all visible contacts one-by-one, then moving to the next zoomed out setting, and so on.
2. **Viewing cues:** When you select a contact, the "Set Contact" button will become active. Click on this button to call up a window that displays all the cues for the contact. Each cue will be labeled and show the particular value for the selected contact.
3. **Threat Assessment:** Once you are finished reviewing the cue information and want to make your threat classification judgment, click on the "Set Threat" button. A small window will appear with two boxes. The red box is labeled "Foe" and the green box is labeled "Friend." Click on the button under the appropriate label. You can change your choice by clicking on the other button. You must select one or the other choice. During the beginning of the training session, you will have to guess. Complete your decision by clicking the "Done" button.
4. **Confidence Judgment:** After indicating whether the contact is Friend or foe, indicate your level of confidence in your decision. You do this by clicking on the confidence bar then sliding the pointer to the left or right to change the confidence value that is displayed to the side. Your confidence rating should indicate the chance, out of a hundred, that you believe your judgment is correct. Once you have set the rating to the appropriate value, click the "Done" button.
5. **Feedback:** After making your confidence judgment, you will be given feedback consisting of the correct threat classification. A window will appear that indicates what classification you gave the contact, whether that classification is correct or incorrect, and the correct threat classification. This feedback is provided to help you learn how to use the cues provided for contacts to make correct threat classifications.
6. **Accuracy and Speed:** We want you to achieve the best performance possible. During the training phase, you should concentrate on learning how to use the cues. We will not look at your overall accuracy or speed in making judgments for the training phase; the feedback is provided solely to help you. During the test phase, however, you will not receive any feedback and you should try to be as accurate as possible in making your threat classification judgments. Being accurate is more important than being fast.

7. **Test Session:** During the first session, you will learn how to distinguish foe from friend contacts through trial-and-error practice. During the second session, we will assess your accuracy in making threat classification judgments. The Test Session will be performed in exactly the same way as the training session but you will not receive any feedback on your judgments.
8. **Describe Friends and Foes:** In the last phase we will ask you to describe the typical friend and typical foe. You will do this by indicating the characteristics (cues) that tended to be associated with friend and foe and estimate the proportion of friends and foes with each of those characteristics. We will give you a sheet listing all the cues to help you in describing friends and foes.

Learning Probabilistic Cues Instructions (Describe Condition)

Preamble

This experiment investigates how people learn to identify what an object is by using different pieces of information (cues). In particular, we want to examine how people deal with cues that are not completely reliable. We are using a naval identification computer task to simulate an identification task aboard a Navy ship. Before we begin the training session, I will describe the task and provide a demonstration of how to operate the simulator.

Instructions

This task may seem complicated at first but we will describe how it works in detail and allow you to practice before beginning the experiment. Please stop me at any time if you have a question.

Let's begin with an overview of the task:

> You are playing the role of an operator in the Operations Room of a naval ship, which is displayed as a blue symbol in the center of your radarscope. Surrounding your ship are asterisks called "contacts." These contacts represent traffic detected by your ship's sensors. Your job is to identify each contact as a friend or foe. Until you identify the contact, its identity is unknown. Your goal in performing the task is to learn the characteristics of friend and foe contacts so that you will be able to describe each of these later in the experiment. Each contact's identity can be determined by several information items that you have access to on your workstation. The information items are called "cues" and come from various sensors on the ship. You will have 4 cues to use in identifying a contact as friend or foe. Friend and foe contacts are different kinds of craft, with different properties that are reflected in different cue values from the sensors. To make your judgments you will first select a contact for identification, review its cues, then indicate whether it is friend or foe. For you to make accurate judgments, you will have to learn the different properties, as reflected in cue values, that describe friend and foe contacts. The first session will be a training session in which you will learn by trial-and-error what cues tend to go with friends and what cues tend to go with foes. It is important to understand that all friends are not exactly alike nor are all foes exactly alike. Due to variations in the exact configurations of the contacts and occasional errors in the sensors, no particular cue will always be 100% associated with either a friend or foe contact. After making your identification for a contact, you will rate how confident you are that your judgment is correct. This procedure will be repeated for each contact.

Some of the information items (cues) will refer to continuous variables, such as speed or altitude, which can take on a wide range of specific values. In this task, however, the exact value of a continuous variable is not important. All continuous variables have a dividing point that cuts the range of values in two. This makes variables such as speed binary – i.e. greater than or less than, higher or lower, above or below, etc. – and all you need to learn is which category is associated with friend and which is associated with foe. For example, a speed greater than 300 nm/hr might be a feature that describes a foe, whereas a friend might be more likely to have a speed less than 300 nm/hr (this is just an example; if speed is a cue in your experiment, don't expect it to behave the way I just said). We have prepared a sheet to indicate what the dividing point is for all continuous variables to help you in this task.

The purpose of the training session is strictly for you to learn how to identify friend and foe. We will not be looking at how many identification judgments you get right or wrong in the training session, so you need not worry about how many contacts you get correct. It is important to bear in mind that, because no cue is 100% associated with either friend or foe, you cannot expect to ever achieve perfect accuracy. That is, it is not necessarily possible for anyone to be correct all the time. We just want you to learn the characteristics of friend and foe so that you will be able to describe a typical friend and a typical foe at the end of this session.

[Be sure that TITAN is loaded and ready to start]

Demonstration

I will now show you the main features of the computer interface and demonstrate how the task is performed.

Main Features

Feature	*Description*
Pointer	The arrow on the screen is a pointer that allows you to view menu items. You will use the mouse to move the pointer around your screen.
Radar Scope	The black circle within the large gray box is the "radar scope," which displays your ship and the contacts within its vicinity.
Ownship	Your ship is located in the center of the shaded circle in the radar scope.
Contact	Surrounding your ship are asterisks called "contacts." A contact is an object detected by your ship's sensors that appears on your radar scope.
Zoom In & Zoom Out	These buttons allow you to magnify and minimize the range of the radar scope. You can zoom in as close as 1 nm and zoom out as far as 1024 nm. These buttons are used to bring an out-of-range contact into view.
Menus	Next to the radar scope are buttons that permit you to inspect cues for the selected contact.
Information Items	These items are the cues for the selected contact and provide information about the contact. **The cues will always appear in a random order for each contact.**
Set Contact	This button allows you to view all the cues for the selected contact at the same time.
Set Threat	This button is used when you are ready to submit your threat assessment. It calls up the Classification Menu.
Classification Menu	This menu is used to indicate the selected contact's threat classification. You make your decision by clicking the appropriate button under the Friend and foe labels.
Confidence Bar	This bar is used to indicate how confident you feel in your classification judgment, on a scale of 0 to 100. You indicate your rating by clicking on the bar and sliding the pointer up or down until the rating shown at the side is the rating you wish to give.
Short and Cumulative	The short average refers to the average error rate of

Averages	the 6 most recent trials. The cumulative average refers to the average error rate across all trials. These averages are provided solely to help you in learning to classify friend and foe.

Step-by-Step Instructions

There are several steps involved in clearing a contact from the display:

9. **Selecting a contact:** First, you must choose a contact. There are no rules governing how you make your selection. You can select randomly, or start at the most zoomed in setting, selecting all visible contacts one-by-one, then moving to the next zoomed out setting, and so on.
10. **Viewing cues:** When you select a contact, the "Set Contact" button will become active. Click on this button to call up a window that displays all the cues for the contact. Each cue will be labeled and show the particular value for the selected contact.
11. **Identification:** Once you are finished reviewing the cue information and want to make your identification judgment, click on the "Set Threat" button. A small window will appear with two boxes. The red box is labeled "Foe" and the green box is labeled "Friend." Click on the button under the appropriate label. You can change your choice by clicking on the other button. You must select one or the other choice. During the beginning of the training session, you will have to guess. Complete your decision by clicking the "Done" button.
12. **Confidence Judgment:** After indicating whether the contact is Friend or foe, indicate your level of confidence in your decision. You do this by clicking on the confidence bar then sliding the pointer to the left or right to change the confidence value that is displayed to the side. Your confidence rating should indicate the chance, out of a hundred, that you believe your judgment is correct. Once you have set the rating to the appropriate value, click the "Done" button.
13. **Feedback:** After making your confidence judgment, you will be given feedback consisting of the correct threat classification. A window will appear that indicates the identification you gave the contact, whether that identification is correct or incorrect, and the true identity. This feedback is provided to help you learn what cues go with friend and what cues go with foe to make correct identification judgments.
14. **Accuracy and Speed:** We want you to be able to accurately describe friends and foes. During the training phase, you should concentrate on learning the likelihoods that a given cue will go with a friend or a foe. We will not look at your overall accuracy or speed in identifying contacts; the feedback is provided solely to help you. During the next phase, however, you will not receive any feedback and you should try to be as accurate as possible in making your threat classification judgments. Being accurate is more important than being fast.

15. **Second Session:** During the first session, you will learn how to describe friend and foe contacts through trial-and-error learning. During the second session, we will have you continue making identification judgments but without feedback. The Second Session will be performed in exactly the same way as the training session but you will not receive any feedback on your judgments. The purpose of the second session is to see how confident you are after the training session.

16. **Describe Friends and Foes:** In the last phase we will ask you to describe the typical friend and typical foe. You will do this by indicating the characteristics (cues) that tended to be associated with friend and foe and estimate the proportion of friends and foes with each of those characteristics. We will give you a sheet listing all the cues to help you in describing friends and foes.

List of symbols/abbreviations/acronyms/initialisms

ANOVA	Analysis of Variance
C2	Command and Control
CF	Canadian Forces
DND	Department of National Defence
MSe	Mean Square Error
NDM	Naturalistic Decision Making
PC	Personal Computer
TITAN	Team and Individual Threat Assessment Network
TTB	Take-the-Best
TTB-C	Take-the-Best-for-Classification

DOCUMENT CONTROL DATA SHEET

1a. PERFORMING AGENCY DRDC Toronto	2. SECURITY CLASSIFICATION UNCLASSIFIED Unlimited distribution -
1b. PUBLISHING AGENCY DRDC Toronto	

3. TITLE

(U) Cue Validity Learning in Threat Classification Judgments

4. AUTHORS

Bryant, David J.

5. DATE OF PUBLICATION	6. NO. OF PAGES
March 12 , 2003	62

7. DESCRIPTIVE NOTES

8. SPONSORING/MONITORING/CONTRACTING/TASKING AGENCY
Sponsoring Agency:
Monitoring Agency:
Contracting Agency :
Tasking Agency:

9. ORIGINATORS DOCUMENT NO.	10. CONTRACT GRANT AND/OR PROJECT NO.	11. OTHER DOCUMENT NOS.
Technical Report 2003-041		

12. DOCUMENT RELEASABILITY

Unlimited distribution

13. DOCUMENT ANNOUNCEMENT

Unlimited announcement

14. ABSTRACT

(U) This report describes an experiment that investigated probabilistic cue learning in a simulated naval warfare threat classification task. The Fast and Frugal Heuristic approach was employed to develop an heuristic, Called the "Take-the-Best-for-Classification" (TTB-C) heuristic, that performs the threat classification task with minimal information and computation. Two variables were manipulated in this experiment. The first, varied between subjects, was the Instruction Set given to participants (Describe vs. Discriminate), which emphasized either the patterns of cue values associated with friend and foe contacts or the differences in typical cue patterns between the two types of contact. The second variable, varied within subjects, was the size of the differences among cue validities (Cue Validity Differences) of the four cues. Four hypotheses were derived from the TTB-C heuristic and tested. Although the results provided support for only one hypothesis, further studies are warranted to explore the potential use of fast and frugal heuristics under conditions of uncertainty, time pressure, and resources costs imposed on data gathering.

(U) Ce rapport décrit une expérience visant à étudier l'apprentissage de repères probabilistes dans une fonction de classification des dangers d'une guerre navale simulée. L'approche heuristique simple et rapide a été utilisée pour élaborer une heuristique, appelée « ne garder que le meilleur en vue de la classification » (TTB-C) qui remplit la fonction de classification des dangers avec un minimum d'information et de calculs. On a manipulé deux variables au cours de cette expérience. La première, qui variait d'un sujet à l'autre, était le jeu d'instructions remis aux participants (Décrire par opposition à Distinguer), qui mettait l'accent soit sur les modèles de valeurs des repères associées aux contacts amis ou ennemis, soit sur les différences entre deux sortes de contact dans les modèles de repères types. La seconde variable, qui variait à l'intérieur des sujets, était l'importance des différences entre les validités des repères (différences de validité des repères) des quatre repères. On a tiré quatre hypothèses de l'heuristique TTB-C et on les a testées. Bien que les résultats n'appuient qu'une hypothèse, il faut faire d'avantage d'études pour explorer l'usage qu'on pourrait faire des heuristiques simples et rapides quand l'incertitude règne, le temps presse et le coût des ressources influe sur la collecte de données.

15. KEYWORDS, DESCRIPTORS or IDENTIFIERS

(U) decision making; cue learning; cue validity; threat assessment; heuristics